

STATISTICS IN BIOLOGY
AND PSYCHOLOGY

SOME RELATED BOOKS

Debajyoti Das
Biochemistry
Biophysics and Biophysical Chemistry

P. K. Giri
Jiban Banerjee
Statistical Tools and Techniques

Saha & Paul
Biostatistics

Pallab Basu
Biochemistry Laboratory Manual

Prakash Chandra Dhara
Computer in Biological Science

STATISTICS IN BIOLOGY AND PSYCHOLOGY

DEBAJYOTI DAS

Reader in Physiology (retired)
Presidency College, Kolkata
Guest-Faculty at Bose Institute, Kolkata
Former Guest Lecturer in Physiology
University of Calcutta
Former Guest Teacher in Human Physiology,
Microbiology, Zoology and Aquaculture
Vidyasagar University

ARATI DAS

Reader in Psychology (retired)
Bethune College, Kolkata
Former Guest Teacher in Human Physiology
Vidyasagar University
Former Counsellor in Psychology (Nursing)
Indira Gandhi National Open University

SIXTH EDITION



ACADEMIC PUBLISHERS

KOLKATA 700 073

E-mail : contact@academicpublishers.in

Website : www.academicpublishers.in

© Copyright reserved by the authors

First published August 1981
Second edition November 1993
Third edition November 1998
Fourth edition July 2003
 Second print January 2004
 Third print July 2005
Fifth edition April 2008
Sixth edition June 2010
 Second print January 2012
 Third print June 2013
 Fourth print August 2014
 Fifth print August 2015

ISBN 978-93-80599-04-5

Price rupees three hundred only.

Published by Bimal Dhur of Academic Publishers, 5A Bhawani Dutta Lane, Kolkata-700073, laser type set by Saugata Ghosh of White Hall Computers, 131B, A. J. C. Bose Road, Kolkata-700014, and printed by Dipankar Dhar at Rajendra Offset and Graphics, 11 Panchanan Ghosh Lane, Kolkata-700009.

ACKNOWLEDGEMENTS

To
Dr. ASOKE GOPAL DATTA
my teacher

The text and figures of the following papers and publications are gratefully acknowledged here.

1. American Cyanamid Company, New York, for their permission to reprint and reproduce Tables I and II of their book *Some Rapid Approximate Statistical Tests* (revised edition, 1944) written by Frank W. Brown and Robert A. Wilson.
2. The Library Board of the University of London, Dr. Frank Yates, FRS, and Longman Group Ltd., London, for their permission to reprint Tables III and IV of their book *Statistical Tables for Biological, Agricultural and Medical Research* (6th edition, 1974).
3. McGraw-Hill Book Company, New York, for their permission to reprint and reproduce (i) Table A of their book *Statistical Analysis in Psychology and Education* (3rd edition, 1971) by George A. Ferguson, where this table has been reproduced from another of their books, *Statistical Methods*, written by J. R. West, and (ii) Table K of their book *Statistical Analysis in Psychology and Education* (3rd edition, 1971) by J. R. Gelfand and Benjamin Fienberg, where this table has been reproduced from another of their books, *Introduction to Statistical Methods* (1971) written by W. J. Dixon and F. J. Massey, Jr.
4. Iowa State University Press, Ames, Iowa, for their permission to reproduce a table of critical F values from their book *Statistical Methods* (6th ed. vol. 1967) written by R. W. Snedecor and William G. Cochran.

ACKNOWLEDGEMENTS

The kind help and gracious cooperation of the following persons and organizations are gratefully acknowledged here.

1. American Cyanamid Company, New York, for their permission to reprint and reproduce Tables 1 and 2 of their book *Some Rapid Approximate Statistical Procedures* (revised edition, 1964) written by Frank Wilcoxon and Roberta A. Wilcox.
2. The Literary Executor of the Late Sir Ronald A. Fisher, FRS, Dr. Frank Yates, FRS, and Longman Group Ltd., London, for their permission to reprint Tables III and IV of their book *Statistical Tables for Biological, Agricultural and Medical Research* (6th edition, 1974).
3. McGraw-Hill Book Company, New York, for their permission to reprint and reproduce (i) Table A of their book *Statistical Analysis in Psychology and Education* (3rd edition, 1971) by George A. Ferguson, where this table has been reproduced from another of their books, *Educational Statistics*, written by J. E. Wert, and (ii) Table K of their book *Fundamental Statistics in Psychology and Education* (6th edition, 1978) by J. P. Guilford and Benjamin Fruchter, where this table has been reproduced from another of their books, *Introduction to Statistical Analysis* (1951) written by W. J. Dixon and F. J. Massey, Jr.
4. Iowa State University Press, Ames, Iowa, for their permission to reproduce a table of critical F values from their book *Statistical Methods* (6th edition, 1967) written by G. W. Snedecore and William G. Cochran.

PREFACE

For three decades now, this book has been proffering the fundamental principles and applications of statistics in the higher studies and scientific investigations of Life Sciences, Psychology and Education. The authors deem this occasion opportune to pay their humble homage to their mentor, the late Rajendra Nath Dhur, but for whose encouragement the first edition of this book would not have seen the light of the day.

The authors persist in the present edition in describing the use of statistics for determining the sample size, drawing samples based on laws of probability, presenting the experimental data in tabulated, graphical and pictorial forms, elucidating the properties of normal, Student's t , binomial and Poisson probability distributions, working out the reliability and validity of tests, testing the significance of the observed results from the probability of correctness of the null hypothesis, estimating errors of inference, analyzing experimental data with numerous parametric and nonparametric tests such as t tests, Wilcoxon's tests, Mann-Whitney test, analysis of variance and Kruskal-Wallis test, using chi square test and G test to find the goodness of fit between the observed distribution and a chosen probability distribution, assessing the linear correlations between variables by Pearson's, Spearman's and Kendall's tests, predicting the probable value of a variable from the values of other variables correlated with the latter, working out item analysis and factor analysis for a psychological test, and standardizing the latter. Mathematical complexities have been confined to what are needed in the current undergraduate and postgraduate syllabi and the subsequent applications of statistics to the experiments in the relevant disciplines, but without any sacrifice of lucidity, clarity, precision and essential methodology. For every statistical method, problems in the form of worked-out sums have been cited as Examples from various relevant disciplines. Different procedures of experimental designs have been systematically elucidated. Basic steps in working out and interpreting several useful statistical tests have been summarized in a chapter in the form of flowcharts for ready reference of the learner. Exhaustive glossaries are given after each chapter for easy access to basic terms and concepts. Numerous sample questions and statistical sums have also been annexed towards the end of this book.

The authors would deem their endeavours rewarded if the intended readers—students of undergraduate, postgraduate and pre-doctoral levels as well as researchers of the relevant disciplines—find this edition more helpful to them.

The authors express their gratitude to Mr. Bimal Kumar Dhur and Mr. Dipankar Dhar for their sincere unhesitant support and unbounded cooperation in publishing this edition.

The authors also convey their appreciation to Mr. Saugata Ghosh of M/s. White Hall Computers and Mr. Biswajit Seal of M/s. Academic Publishers for their roles in the publication work.

Kolkata,
Bengali New Year Day,
15 April 2010.

Debajyoti Das
Arati Das

THANKSGIVING

The authors gratefully thank the following academic personalities for the encouragement, advice and cooperation received from them :

Prof. H. Y. Mohan Ram, formerly of the University of Delhi ; Prof. Ananga Mohan Chandra, Prof. Tushar Kanti Ghosh, Prof. Mallika Banerjee, Prof. Pritha Mukherjee and Dr. Sadhan Dasgupta of the University of Calcutta ; Prof. Prakash Dhara, Prof. Debidas Ghosh and Dr. Sujata Maity of Vidyasagar University ; Dr. Gautam Pal of Kalyani University ; Dr. Devashis Sen and Dr. Subrata Ghosh of Presidency College ; Dr. Arnab Chaudhury of City College ; Dr. Bitan Bindu Banerji of Uttarpara Raja Peary Mohan College ; Dr. Dilip Kumar Nandy and Dr. Partha Pratim Chakraborty of Raja N. L. Khan Mahila Mahavidyalaya ; Dr. Shyamali Sen of Bethune College ; Dr. Asima Das and Dr. Sharadiya Ray of Serampore College ; Dr. Bhaswar Mukherjee of Surendranath College ; Dr. Sahana Sen of Rammohun College ; Dr. Leena Nair Sengupta and Ms. Sumita Mukherji of Gokhale Memorial Girls' College ; Dr. Snigdha Mitra and Ms. Sujata Chatterji of Bijaykrishna Girls' College ; Dr. Purushottam Pramanik of Hooghly Mohsin College ; Prof. Gouri Sankar Sa and Prof. Tanya Das of the Bose Institute.

CONTENTS

1. Statistics, variables and samples 1-14

Applications of statistics 1 ; Variables 2 ; Population and sample 6 ; Sampling methods 7 ; Parameter and Statistic 11 ; Glossary 12.

2. Presentation of data 15-40

Frequency distributions 15 ; Pie diagram 20 ; Bar diagram 20 ; Frequency polygon 22 ; Histogram 26 ; Frequency distribution curve 30 ; Ogives 30 ; Scattergram 33 ; Linear graph 34 ; Exponential curve 37 ; Semi-logarithmic graph 38 ; Glossary 39.

3. Statistics of location 41-55

Classification 41 ; Mean 41 ; Geometric mean 45 ; Quantiles or fractiles 46 ; Percentile ranks 48 ; Median 49 ; Mode 53 ; Glossary 55.

4. Statistics of dispersion 56-72

Measures of dispersion 56 ; Range 56 ; Mean deviation 57 ; Standard deviation 57 ; Variance 64 ; Central moments 67 ; Quartile deviation 67 ; Coefficient of variation 70 ; Coefficient of quartile deviation 71 ; Coefficient of dispersion 71 ; Glossary 72.

5. Sampling statistics 73-83

Sampling errors 73 ; Sampling distributions 73 ; Standard errors 74 ; Standard scores 80 ; Degrees of freedom 82 ; Glossary 82.

6. Probability distributions 84-109

Probability 84 ; Probability distributions 85 ; Normal distribution 86 ; Best-fitting normal distribution 89 ; Skewness 92 ; Kurtosis 94 ; Student's t distribution 95 ; Confidence intervals 97 ; Binomial distribution 100 ; Poisson distribution 105 ; Glossary 108.

7. Testing of hypothesis 110-143

Difference between means 110 ; Null hypothesis 111 ; Levels of significance 112 ; Errors of inference 114 ; One-tail and two-tail tests 115 ; Significance of difference between means using z scores 117 ; t tests for significance of difference between means 123 ; Glossary 143.

8. Correlation and regression 144-207

Correlation 144 ; Product-moment correlation 145 ; Partial correlation 162 ; Multiple correlation 164 ; Spearman's rank correlation 167 ; Kendall's rank correlation 172 ; Point biserial r 179 ; Biserial r 183 ; Yule's phi coefficient 187 ; Tetrachoric r 189 ; Contingency coefficient 191 ; Regression 193 ; Multiple regression 203 ; Glossary 206.

9. Nonparametric statistics 208-246

Nonparametric statistics 208 ; Chi square tests 208 ; G test 222 ; Wilcoxon signed rank test 227 ; Wilcoxon composite rank test or rank sum test 231 ; Mann-Whitney U test 237 ; Median test 242 ; Glossary 245.

10. Psychological test construction 247-280

Psychological variables 247 ; Principles of test construction 249 ; Power and speed tests 250 ; Item analysis 250 ; Reliability of a test 257 ; Validity of a test 263 ; Expectancy table 268 ; Factor analysis 268 ; Standardization and norms 270 ; Glossary 278.

11. Analysis of variance 281-328

Experimental design 281 ; Analysis of variance 288 ; One-way anova 291 ; Multiple comparison tests 300 ; Kruskal-Wallis nonparametric anova 310 ; Two-way anova 315 ; Glossary 326.

Steps in statistical tests 329-341

A significance test 329 ; t test for $(\bar{X}_1 - \bar{X}_2)$ of large independent groups 330 ; t test for $(\bar{X}_1 - \bar{X}_2)$ of small independent groups 331 ; t test for $(\bar{X}_1 - \bar{X}_2)$ of small single group 332 ; Using z score for $(\bar{X}_1 - \bar{X}_2)$ of large independent groups 333 ; Mann-Whitney U test for $(\bar{X}_1 - \bar{X}_2)$ 334 ; Chi square test for goodness of fit 335 ; Product-moment correlation using sum of products 336 ; Product-moment correlation using raw scores 337 ; Simple linear regression of Y on X using sum of products and sum of squares 338 ; Simple linear regression of Y on X using raw scores 339 ; Kruskal-Wallis nonparametric anova 340 ; One-way anova 341.

Bibliography 342-343

Sample questions 344-373

Appendix : Tables 374-384

Index 385-387

1. STATISTICS, VARIABLES AND SAMPLES

Each experiment or test explores a property or event which varies from individual to individual in a group, or from time to time in the same individual. Dependable inferences can seldom be drawn by a mere straightway inspection of the experimental observations. So, the experimentally obtained data need statistical treatment for proper inferences. Statistics is also used in the scientific designing of experiments and tests.

1.1 APPLICATIONS OF STATISTICS

Statistics is the science of the methodology for the scientific collection, systematic presentation, mathematical analysis and interpretation of the data, and for drawing inferences about the explored property or phenomenon in the relevant population. In this respect, statistics has the following basic applications.

(a) Statistics is used in *designing experiments* scientifically and *minimizing experimental errors*.

(b) It is used in *drawing a representative sample* from the population for an experiment or survey, and in *fixing the sample size* for dependable inferences.

(c) Statistics is used for the systematic arrangement of the observed data to express them in common communicable forms. It thus provides for a *meaningful description or presentation* of the studied property or phenomenon.

(d) There is always a probability that the observed results of an experiment, using a sample instead of the entire population, occurred by chance in the particular sample and would not occur if the entire population were studied. Statistics is used in assessing this *probability of chance occurrence* of the observed results — the inference is then drawn according as this

probability is too low or too high. Thus, statistics helps to *generalize the findings of a sample to the entire population*.

(e) Statistics also estimates the *probability of errors* in the inference.

(f) Statistics is used in estimating the *reliability* of an experimental method or test in measuring the specific relevant property, through a mathematical assessment of the consistency of results on repeating the test on the same sample.

(g) It is also used in estimating the *validity* of a test, i.e., the capacity of that test to measure the intended property in exclusion of other closely related ones.

(h) Statistics explores and quantifies the *magnitude and direction of relation* between the changes of two or more properties, events or phenomena, as well as their *cause and effect relations*.

(i) It may be applied to *predict* mathematically the most probable value of a property or event in an individual or case, from the observed value(s) of one or more of other related properties or events in that individual or case.

(j) The causative factors for the variations of a property or event are explored by the statistical method of *analysis of variances*.

In such various ways, statistics is used to study numerous aspects of human problems like individual differences in physicochemical properties and psychological characteristics, educational, epidemiological, sephological, ergonomic, industrial and population problems, pollution hazards, market and employment surveys, medical therapy and scientific research.

1.2 VARIABLES

A *variable* is any property, phenomenon or event that varies in quantity or quality either *spatially*, i.e., from case to case, individual to individual, place to place, organ to organ, and cell to cell even at the same instant, or *temporally*, i.e., from time to time even in the same case, in the same individual, in the same tissue, in the same cell or at the same location. Almost all properties and constituents of a living being are variables; statistics is mostly concerned with the study of such variables. For example, even when measured at the same time, the systolic arterial pressure may differ in different arteries of a person; the trunk length, wing length or tracheal ventilation volume differs from insect to insect of the same species; the interorbital width differs from bird to bird of a species; the cell count differs from culture to culture of the same micro-organism; the blood sugar, height, weight, blood volume, surface area, vital capacity, skinfold thickness, urine volume, oral temperature, memory, learning ability, intelligence, job performance, lens power and O_2 consumption differ from human to human even at the same moment; the animals of a species differ in sex, age, phenotype, genotype, litter size, femur length, coat color and ferocity; atmospheric PO_2 differs at different altitudes at the same moment while the sea water temperature differs at the same time from place to place. All such are *spatial variations*. On the contrary, the blood sugar, oral temperature, blood pressure, urinary NPN concentration, blood volume, sweating rate, height, body weight, body surface area, alertness, motivation level, O_2 consumption and urine volume may vary in the same person at different times; the trunk length, wing length or tracheal ventilation changes in the same insect with age; the femur length, interorbital width and coat color change in the same animal as it grows. All such are examples of *temporal variations*.

Classes of variables

1. Measurement variables :

Magnitudes or amounts of these variables can be measured or counted, and expressed in numerical values. Examples : age, height, weight, litter size, femur length, wing length, interorbital width, heart rate, respiratory rate, blood volume, blood pressure, hemoglobin concentration, blood sugar, urinary NPN, bacterial or blood cell counts, reaction time, nerve impulse velocity, body temperature, vital capacity, body surface area, respiratory CO_2 output, and cell size. Numerical values of a measurement variable are called its *scores*. The scores of a measurement variable show clearly : (i) whether two such scores are *different from or equal to* one another, (ii) whether one of them is *higher or lower* than the other, and (iii) *by how much* one score exceeds or falls short of the other. For example, if two persons have pulse rates of 100 and 72 per minute respectively, their pulse rates differ, the former being higher than the latter by as much as 28 beats per minute. Measurement variables belong to the following two sub-classes.

(a) *Continuous variables* : These can be measured and expressed not only in whole units, but also in any infinitely small fraction of such units. So, continuous variables can have an infinite number of scores even between two whole units — their scores can increase or decrease by infinitely small fractions of whole units, with no limit to the smallness of those fractions. So, the possible scores of such a variable form a *continuous series with no intervening gaps*. For example, the height of a human or the wing length of an insect can be measured upto the level of centimetres, millimetres, nanometres and still lower, depending upon the degree of fineness of the measuring instrument and the precision of measurement. Thus, a continuous variable should yield *continuous metric data with no real gaps*; however, such data from an actual study show apparent gaps owing to (i) limitations of fineness

and precision of measurement, and (ii) accidental absence of individuals having the missing intervening scores in the sample. Examples : age, height, weight, surface area, blood or urine volume, blood sugar, serum cholesterol, body temperature, vital capacity, pH, impulse velocity, reaction time, cell size, blood pressure, alveolar Po_2 , organelle dimensions, microbial fermentation rates, and the scale scores of intelligence, anxiety, interest, aptitude, achievement, etc.

(b) *Discontinuous, discrete or meristic variables* : These can have only certain specific values, and no intermediate value between those fixed ones. Their scores can be only in whole units in most cases, and upto certain fixed fractions of the whole unit in other cases, because of the impossibility or impracticability of further subdivision of their measures. So, their scores may increase or decrease by that fixed amount only, and not by any smaller fraction of the unit. Their scores thus form *discontinuous data* with *real gaps* between consecutive scores. Examples include the *enumeration data* from the counting of individuals or cases, such as litter size, family size, number of animal digits, heart rate, respiratory rate and impulse frequency — such counts are limited to whole numbers of cases and cannot be reasonably extended to fractions of cells, digits, litter mates, respiratory cycles, etc. Heart rate, however, is often treated as a continuous variable during statistical analysis.

There are two types of scales for measurement variables. (i) *Interval scale* : Some measurement variables such as temperature, intelligence, memory and learning ability cannot be absolutely absent — with a zero value — anywhere or in any case ; so, the scale used for such a variable has *no true zero point* — each such *interval scale* may start from any conveniently chosen *arbitrary zero point* ; e.g., celsius, fahrenheit and kelvin scales for temperature, having different arbitrary zero points. In the absence of a true zero point, two scores in the interval scale cannot be multiplied or divided by each other and a ratio of the two

cannot be directly worked out. (ii) *Ratio scale* : Many measurement variables such as height or length, mass, weight, volume, pitch and loudness can be totally absent so as to have a zero value in some cases or at some places ; so, the scale for such a variable has a *true zero point* — all such *ratio scales* for a particular variable (e.g., both centimetre and inch scales for length) start from the same true zero point. The existence of a true zero point enables the scores in a ratio scale to be multiplied or divided by each other and to be used for working out their ratios. Ratio scales are far more extensively used in life sciences than in psychology ; however, ratios between stimuli in psychophysics, calorie expenditures in various job activities, and times taken in assembly tests in industrial psychology are worked out using ratio scales.

2. Ordinal or ranked variables :

These consist of variables like attentiveness, leadership quality, social adjustment, trustworthiness, ferocity and docility, which vary distinctly in magnitude or intensity among members of a population, but their magnitudes cannot be measured quantitatively. Individuals of a sample may be graded into ranks in ascending or descending orders according to the *relative magnitudes* of the relevant variable in them ; but these ranks give neither any absolute quantitative measure of the variable in an individual nor the magnitude of differences between two individuals. Thus, the ranks of an ordinal variable show (i) whether two individuals are *different from or equal to* one another and (ii) whether one is *higher or lower* than the other with respect to that variable, but (iii) cannot indicate *by how much* one is higher or lower than the other, and (iv) gives no indication about whether the *difference in magnitude* of the variable between any two consecutive ranks equals, exceeds or falls short of the difference between two other consecutive ranks. Thus, the difference in ferocity between animals ranked 1 and 2 may very well be either equal to, or higher than or lower than the difference between those

ranked 2 and 3. Moreover, the set of ranks in a sample, each given in whole number only, constitutes a *discontinuous series* with *real intervening gaps*, and thus constitutes a *discontinuous variable*.

3. Nominal variables or attributes :

Some variables such as sex, race, caste, religion, profession, species, blood groups, mother tongue, eye color, skin color, hair color, living-and-dead, pregnant-nonpregnant, success-failure, HIV positive-negative, and some psychological stimuli cannot be measured, quantitatively nor expressed in numerical scores. Nor can the individuals of a sample be ranked about that variable because of no discernible difference in magnitude of the variable. Such *attributes* or *nominal variables* can only be studied qualitatively. According to the qualitative similarity or difference of the members of a population with respect to such a variable, the latter is divided into classes with intervening gaps — an attribute, thus divided into only two classes, is called a *dichotomous variable* ; e.g., sex, pregnant-nonpregnant, success-failure. Individuals of each such class are qualitatively similar to each other and differ from those of other classes, but these differences cannot be assessed quantitatively. So, it can be stated (i) whether two individuals are *similar to or different from* one another about such a variable, but (ii) not whether one is *higher or lower* than the other. Although such a variable cannot be measured to give numerical scores for the respective individuals of a sample, the numbers of individuals or cases counted in its different classes, form *enumeration data* for statistical treatment (Table 2.1)

4. Derived variables :

Each such variable consists of the variable relations between the directly measured/counted scores of two or more measurement variables, expressed as their ratios, proportions or percentages. Examples : respiratory quotient, BMR, color index, cephalic index, hematocrit

value, renal clearance, clearance ratio, digestibility coefficient, ratio IQ, educational quotient, etc.

Variables in experiments

An experiment is undertaken to study the changes of a particular variable (dependent variable) in a sample on exposure of the latter to other given variables (independent variables). According to their roles, variables involved in the experiment are categorized into dependent, independent and relevant variables, each *belonging to any of the classes* described earlier.

1. Dependent variable :

The variable, whose possible change due to exposure of the sample to some other variable(s) is being studied in an experiment, is called the *dependent variable* in that experiment — its changes, if any, may be supposed to depend on the effects of the other (independent) variable(s) applied on the sample. Thus, while studying the changes in blood sugar after insulin administration to a sample of animals, the blood sugar concentration whose changes are measured constitutes the dependent variable ; in an experiment on changes in respiratory rate after ephedrine administration to a sample, the respiratory rates constitute the dependent variable. It may be noted that in the first example, the dependent variable, viz., blood sugar, is a continuous measurement variable ; but in the second instance, respiratory rate is a dependent variable of the discontinuous type. Again, in an experiment on the effect of practice on the rankings of workers in a given job performance, the ranks in the job performance constitute the dependent variable, belonging to the ordinal class. Although changes of the dependent variable are explored or measured in the experiment, this variable is *not controlled* or “fixed” by the investigator and is thus *liable to random errors*. In each experiment, there is a *single dependent variable* ; an investigation into the effects of an applied independent variable (say, epinephrine) on more than one dependent variable (say, blood

sugar and blood lactate) should be considered as a combination of more than one experiment, done simultaneously with the same sample using an identical independent variable, but each having a single dependent variable (viz., blood sugar and blood lactate, respectively).

In psychology, any behavioural variable, whose changes are assessed in a test or experiment, constitutes the dependent variable and is generally called the *criterion* in that test. For example, in a test for degrees of steadiness in a sample of humans after different levels of practice, steadiness is the dependent variable or criterion while practice is the independent variable. Similarly, in psychological tests for assessing job satisfaction at different levels of attitude of management, job satisfaction is the criterion while management attitude is the independent variable.

2. Independent variable :

The variable, whose effects on the dependent variable are studied in an experiment, is called the *independent variable* in that experiment — the changes of the independent variable are not dependent on the changes of the dependent variable in the relevant experiment. In an experiment, the dependent variable is always a single one, but there may be *one or more independent variables* in an experiment and each of them may *belong to any of the classes* of variables described earlier. In an experiment on the effect of insulin administration on blood sugar, doses of insulin constitute the lone independent variable of the continuous variable class ; in a study of the difference in wing lengths between males and females of an insect species, sex is the single independent variable of the nominal variable class ; in a study of the combined effects of androgenic steroid administration and sex on athletic performance, the doses of the androgenic steroid constitute a continuous measurement variable while sex is a nominal variable, both being the independent variables in this experiment.

In psychology, the independent variable whose effect on a given behavioural variable (criterion) is explored in a test is called the *predictor*, supposing that the criterion score of an individual may be predicted from the quantity/nature of the independent variable affecting him.

Independent variables may belong to any of the following types, according as they are under or beyond the control of the investigator.

(a) *Treatment variables* : These are extensively controlled by the investigator. Their qualitative forms, quantitative amounts or applied doses as well as methods of their application are either fixed or altered by the investigator in deliberate, predetermined and precise manners in an experiment to study the effects of such changes on the dependent variable ; they are not otherwise allowed to change, vary or fluctuate at random, while in use in the experiment. They thus *do not suffer from random errors* during the experiment. Examples : applied doses of a hormone, drug, pesticide or radiation ; extirpation or transplantation of an organ ; pitch or loudness of a stimulating sound ; periods of practice ; voltage or amplitude of stimulating current ; intensity or wavelength of light stimulus.

(b) *Classification variables* : They are not directly controlled or “fixed” by deliberate manipulations by the investigator — rather, they change at random owing to their inherent nature and external factors, all beyond the control of the investigator. So, they are *liable to random errors*. Often, they have been existing and affecting the dependent variable since long before the experiment. In an experiment, the sample includes cases either already exposed to different levels or already belonging to different classes of the given classification variable, and is subjected to the assessment of the intended dependent variable. Examples : sex, age, breed, genetic disorders, cosmic rays, atmospheric pressure or temperature, phenotypes, race, blood groups, intelligence levels, temperament and personality.

In the cited experiment about the effect of insulin administration on blood sugar, the independent variable is a treatment variable constituted by the "fixed" doses of insulin ; in a study of the difference in wing length between male and female insects, sex is a classification variable beyond the control of the investigator ; in a study of combined effects of androgenic steroid administration and sex on athletic performance, "fixed" doses of the administered steroid constitute a treatment variable while sex is a classification variable.

• 3. Relevant variables :

In every experiment or test, the dependent variable may be affected by numerous variables that are not intended to be used by the investigator as independent variables (predictors) to study their effects on the dependent variable (criterion), but may still affect the latter if not effectively controlled by the investigator. Such variables are called the *relevant variables* for the given experiment or test. They may be classified as follows.

(a) *Subject-relevant variables* : These arise from variations of numerous properties or qualities of the subjects (individuals) chosen for the sample. In life sciences, subject-relevant variables include age, sex, race, strain, breed, body weight, genotype, phenotype, etc., of the subjects. In psychology, such variables include aptitude, interest, motivation, personality, intelligence, etc. of the subjects.

(b) *Situational relevant variables* : These arise from experimental situations and physical or social environments. In life sciences, they include variables in laboratory experiments such as pH, ionic strength, temperature, density and equipment design, and noise, distraction, humidity, workshop space and ventilation in field or workshop studies. In industrial psychology, such variables are called *organizational and social variables*, and include incentives, training, supervision and attitude of management.

(c) *Sequence-relevant variables* : These arise from the sequence or order of application of independent variables. In life sciences, they include the order of application of different doses or levels of drugs and other treatments on the subjects. In psychology, they include practice, fatigue and monotony affecting a criterion such as performance.

The unwanted effects of relevant variables are sought to be minimized by experimental designs such as randomizations of sampling and of application sequence of independent variable, use of equivalent or matched groups of subjects for different doses or levels of application of independent variable, and other counterbalancing techniques.

1.3 POPULATION AND SAMPLE

Population

The vast group or aggregate of all animate or inanimate individuals, cases or events that possess some form or amount of the variable under investigation in an experiment, test or survey, constitutes the *population* for the latter. It is of the following two types.

(a) *Infinite population* : Such a population is so vast and widespread that all its members cannot be counted and its size cannot be precisely determined ; e.g., population of all diabetic humans, all spawning salmons, all sickle cell anemia cases, all thalassemia patients, all Down syndrome cases, cholinergic neurons of all humans, all hypophysectomized white rats of a particular strain, all locusts exposed to a pesticide, and all fluoride-treated cultures of *lactobacillus casei* — in the whole universe.

(b) *Finite population* : Such a population is so limited in size and location that all its members may be precisely counted ; e.g., populations of all examinees in the secondary school-leaving examination, all athletes in an Olympiad or Asiad, all Mansarovar pilgrims of a particular year, all victims of the Bhopal gas accident, all donors in

a city blood bank, and all crocodiles in a sanctuary.

Measures like mean, median and standard deviation of any variable may be considered to remain more or less constant in a population, whether finite or infinite, because of its large size. But it is too laborious, expensive and impracticable to conduct any experiment or survey with the entire vast population ; even if so attempted, some of its members may be accidentally left out of the study without the knowledge of the investigator so that such omissions can be neither compensated nor estimated for the consequent errors.

Sample

A *sample* is a relatively small group of members (individuals, cases, events, etc.) chosen from a population for being subjected to an experiment, test or survey instead of that population. Use of a sample in an experiment or survey is the only practical solution of problems, cited above, in using the entire population in such work. Because observations recorded from a sample are to be used for a generalized inference about the entire population, the sample should be drawn by well-planned scientific methods to make it truly *representative* of the population. Such a representative sample should fulfil the following criteria.

(a) Proportions of different types of individuals or events in the sample should conform to their respective proportions in the population.

(b) Variations of the dependent variable among the members of the sample should not exceed the range predictable by the laws of probability from the corresponding variations in the population.

(c) Different samples drawn from the population should have fairly close summary values (statistics) such as the mean so that it is reasonable to assume that the arithmetic average of any such statistic of different samples would coincide with the corresponding summary value (parameter) of the population.

1.4 SAMPLING METHODS

A sample should be fairly *representative* of the population, because (i) the observations made with the sample have to be used for *generalized inferences* about the entire population, and (ii) the *errors of inference*, arising from the planned exclusion of the rest of the population from the study, have to be estimated. So, the sample is to be drawn by well-planned methods, described below.

1. Judgement sampling

In this method, the investigator deliberately chooses some members of the population for forming the sample, *judging* those particular members as the representatives of the population, often depending arbitrarily on some quality or characteristics in them. No random sampling is used, depending on the laws of probability, for drawing the sample. Such a sample is frequently biased and *devoid of any representative nature*, because of (i) personal judgement in the choice and (ii) deliberate unplanned exclusion of other members of the population. So, such judgement sampling should be avoided.

2. Probability sampling

Probability sampling is done by choosing *at random* the individuals or cases from a population for inclusion in the sample, depending (i) on the *laws of probability* and (ii) on the *frequency of each type of individual* or case in the population, (iii) leaving *no scope for arbitrary choice* by any person. This ensures that (i) individuals or cases of a larger section of the population would get chosen for the sample in proportionately higher number than those of a smaller section, and (ii) individuals or cases of each section or class of the population would be included in the sample in the same proportion as that in the population. Methods of probability sampling are as follows :

(a) Simple random sampling :

This consists of the random choice of individuals or cases for a sample, depending on

the laws of probability, from the undivided whole of a *small, finite and homogeneous population*. Simple random sampling ensures that (i) no element of conscious or unconscious bias, whim or personal factor of the investigator affects the choice, (ii) each member of the population has an equal probability of being chosen for the sample, and (iii) the choice of any member of the population for the sample is independent of the choice or exclusion of any other member. The *sample size*, i.e., the number of individuals or cases in the sample, should be determined statistically because (i) the smaller the sample, the higher is the probability of exclusion of particularly the extreme and rare cases of the population, and (ii) a minimum sample size is required for reducing the errors due to the planned exclusion of the rest of the population.

In the unsophisticated *card drawal method*, all members of the population are assigned serial numbers which are written singly on separate cards. The cards are shuffled and mixed together, and as many of them are then blindly picked up as the required sample size. The individuals bearing the serial numbers on the drawn cards are included in the sample. There are the following two ways of random sampling.

(i) In random sampling *with replacement*, a member once chosen for the sample continues to be considered for all subsequent choices also. In the card drawal method, for example, a card once drawn would be replaced in the bunch of cards after noting its serial number; this would give a chance to the same card for being drawn again. This procedure keeps the probability of choice of each individual identical and unchanged at any step of choice, because all the choices continue to be made from the unchanged total number of individuals. Thus, for each drawing from a population having a total of N number of members, every member has the same probability of $1/N$ of being chosen for the sample. But it is impracticable to work with a sample in which an individual gets included more than once due to the

replacement.

(ii) In random sampling *without replacement*, an individual once chosen is kept out of consideration for all subsequent choices — a card once drawn is not replaced in the bunch of cards. Of course, the probability of being chosen for the sample rises from choice to choice instead of remaining identical, because with each choice the total number of individuals left for the next choice falls by 1. Thus, for the r th choice from a population of size N , the members left for that choice would number $(N - r + 1)$, each with a probability of $1/(N - r + 1)$ for being chosen for the sample; but for the next or $(r + 1)$ th choice, this probability would rise to $1/(N - r)$. This progressive rise in probability is, however, negligibly low or practically absent for small samples from infinite ($N = \infty$) or large finite populations.

Random number method: A more scientific and precise method of random sampling is based on a *random number table* of sequences of a large number of digits compiled and arranged at random with a random number generator. For *sampling without replacement*, all members of the population are assigned identity numbers serially, each such number comprising as many digits as those in the population size. If, for example, a sample of 11 individuals has to be drawn from among a total of 48, the identity numbers should be in two digits. However, many random numbers will have to be rejected before the sample size is fulfilled if only one identity number is allotted to each member; so, after assigning the first series of identity numbers — say, 01 to 48 — to the members, the allotment of identity numbers is continued in the same order — say, starting from 49 and going upto 96 (*Example 1.4.1*). Thus, the members bearing the identity numbers of 08 and 11 respectively in the first series, bear also the second-series numbers of respectively 56 and 59; the remainder left after dividing any second-series number (say, 59) by the last first-series number (48 here), equals the first-series identity number

(viz., 11 here) of the same individual. Next, any digit of any set of random number of the table is chosen at random ; passing either horizontally along the row or vertically along the column from this chosen digit and omitting none of the digits in this path, each set of as many consecutive digits as the digits in the population size is noted as a chosen random number and used as follows : (i) when a chosen random number coincides with a first-series identity number, the individual bearing that identity number is included in the sample ; (ii) when the chosen random number coincides with a second-series identity number, that number is replaced by the corresponding first-series identity number and the individual bearing the latter is included in the sample ; in the aforementioned *Example 1.4.1*, if a random number of 76 gets chosen, the individual bearing the identical second-series identity number of 76 and thus the corresponding first-series identity number of 28 gets included in the sample ; (iii) any chosen random number higher than the last second-series identity number (96 here) is rejected ; (iv) random number of 00, if chosen, is also rejected ; (v) in case of sampling without replacement, as in the given example, any identity number getting chosen a second time is also rejected. These procedures are continued till the required sample size is reached.

(b) *Stratified random or quota sampling :*

This method is used if the *population is large and heterogeneous* with distinct subpopulations or strata, each differing from the others with respect to a property related to the variable under study, but being practically homogeneous in itself. Here, simple random sampling is applied *separately on each stratum* to draw that many individuals from it as would correspond to the proportional size of that stratum in the population — the proportions of individuals in different strata of the sample would thus correspond to those in the respective strata of the population. For instance, for a sample of 160 adult humans to be drawn from a population having a male : female

ratio of 55 : 45, a total of 0.55×160 or 88 males and 0.45×160 or 72 females should be chosen by separate simple random samplings from the respective strata of the population, ensuring a male : female ratio of 55 : 45 in the sample also. Evidently, while all members of a particular stratum have the same probability of being chosen for the sample, this probability varies from stratum to stratum according to their proportional sizes in the population. Stratified random sampling minimizes the errors due to the heterogeneity of the population and thereby increases the representative character of the sample.

(c) *Multistage sampling :*

This method is used when the *population is vast and widespread*. Utilizing some existing steps of the population, it is divided into a number of successive stages ranging from the entire population to the individuals. Simple random sampling is then applied at each stage separately. For example, for an experiment or test about an anthropometric variable or a learning ability of urban class X students of a state, a few of the towns may first be chosen at random as the *first-stage units* from amongst all the towns ; a number of schools are next chosen at random in the chosen towns as the *second-stage units* ; then a specific number of class X students are chosen at random from the chosen schools at the *third and final stage* to constitute the sample.

(d) *Fixed interval sampling :*

Where all members of a population have already been arranged and numbered serially in alphabetical, chronological, registration- or performance-based or some other systematic order, the first member for the sample may be chosen purely at random. This may be followed by including in the sample other members occurring at *randomly predetermined regular intervals* along the existing serial order. Thus, such a *systematic or fixed interval sample* of students may be drawn from a class by choosing at random, say, Roll No. 4 as the first individual and then

drawing, say, every subsequent seventh Roll no. Though apparently a random method, it may yield a biased sample if the initial serial order has been based on some property related to the variable to be investigated.

(e) *Purposive sampling :*

In this method, random sampling is done from

only a limited section of the population, considered by the investigator to represent the entire population faithfully with respect to the variable under study. But such a sample may be biased if the assumption about the representative nature does not hold good for the relevant section of population.

Example 1.4.1.

Draw a sample of 11 scores by random sampling without replacement, from the following 48 body weight (kg) scores, using the random numbers provided below the data.

Scores :

68, 59, 57, 64, 52, 60, 62, 57, 60, 62, 57, 61, 63, 47, 55, 50, 59, 71, 70, 49, 56, 67, 82, 56, 45, 73, 66, 46, 72, 44, 58, 62, 53, 78, 40, 43, 60, 54, 76, 81, 38, 77, 41, 47, 66, 65, 72, 70.

Random numbers :

4227	1701	7616	3200	7336	3974	7025
4108	8414	8116	6688	7061	0344	4348

Solution :

(a) As the total number of scores has two digits ($n = 48$), identity numbers of two digits are allotted to the scores. Proceeding serially from the first to the last score of the data, all the scores are first allotted one identity number each (viz., 01 to 48); then, the allotment of identity numbers (viz., 49 to 96) is continued a second time starting again from the first to the last score.

Table 1.1. Allotment of identity numbers (IN) to the scores.

Scores :	68	59	57	64	52	60	62	57	60	62	57	61	63	47	55	50	59	71	70	49	56	67	82	56
IN-1 :	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
IN-2 :	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
Scores :	45	73	66	46	72	44	58	62	53	78	40	43	60	54	76	81	38	77	41	47	66	65	72	70
IN-1 :	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
IN-2 :	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96

(b) Any digit of any set of the provided random numbers is then chosen at random to start with – here the first digit (viz., 1) of the second set (viz., 1701) of the first row is so chosen. Starting from this, groups of each two consecutive digits are recorded from the random number sets proceeding serially along their rows. A recorded random number is rejected if (i) it is higher than the highest identity number (viz., 96) of the second series, (ii) the random number is 00, and (iii) the random number has already been recorded once earlier. Thus, the recorded random numbers are :

17, 01, 76, 16, 32, 00 (rejected), 73, 36, 39, 74, 70, 25, 41, 08, 84, 14, 81, 16 (rejected), 66,

(c) The recorded random numbers are replaced by the corresponding identity numbers (IN) of Table 1.1. If a random number coincides with a first-series IN, the latter replaces that random number ; if a random number coincides with a second-series IN, the first-series IN corresponding to the latter replaces that random number ; if a random number is higher than the total number of scores (viz., 48), it is then replaced by the corresponding IN

of the first series ; identity numbers (viz., 25 and 36) which have already come once earlier, should also be rejected on the second occasion as the sampling is being done here without replacement. Thus, the final identity numbers chosen are as follows :

17, 01, 28, 16, 32, 25, 36, 39, 26, 22, 25 (rejected), 41, 08, 36 (rejected), 14, 33, 18,

(d) The sample size being 11, the scores bearing the first 11 chosen identity numbers are included in the sample and given below. (If any score would have occurred a second time, it would have been rejected.)

59, 68, 46, 50, 62, 45, 43, 76, 73, 67, 38.

1.5 PARAMETER AND STATISTIC

Parameter

A parameter is a summary value or numerical index like mean, standard deviation, median or variance of a variable for the entire population. As an entire population is not usually subjected to experimental methods, a parameter is not directly worked out or precisely known in most cases. However, measures like mean and median of a sample are frequently straightway used as the estimates of the corresponding parameters of the population from which the sample has been drawn (*point estimates*). Moreover, a narrow range of scores, called a *confidence interval*, is sometimes worked out from the measures like mean and standard deviation of the scores of a sample with a stipulated probability of the parameter falling within that interval (*interval estimates*). A parameter is not expected to change so long as the population remains unaltered.

Statistic

A statistic (plural : statistics) is a summary value or numerical index like mean, median, standard deviation or variance of the scores of a variable in a sample. It can be *directly worked out* from the measured scores of the variable in the sample, and can be used either directly as a point estimate of the corresponding population parameter, or in working out a confidence interval as an interval estimate of the latter. A statistic *varies from sample to sample* of the same population due to different individual scores of different samples. A statistic frequently differs

from the corresponding parameter, such a difference being known as the *sampling error*. The values of a statistic for different samples from the same population are scattered around the value of the corresponding population parameter to form a *sampling distribution* of that statistic. Statistics may be classified as follows.

Table 1.2. Symbols for a few statistics and parameters.

Measure	Statistic	Parameter
Mean	\bar{X}	μ
Standard deviation	s, s_X	σ, σ_X
Variance	s^2	σ^2
Correlation coefficient	r	ρ
Standard error of mean	$s_{\bar{X}}$	$\sigma_{\bar{X}}$

(a) Descriptive statistics :

These statistics describe the properties of a sample with respect to the given variable or variables. They include mean, median, mode, percentiles, standard deviation, variance, coefficient of variation, correlation coefficient, etc. They belong to the following three categories according to the property of the sample described by them.

(i) *Statistics of location* : They give the location of either a central position or some other specific point of the frequency distribution of scores in a sample on the scale of the variable. Statistics of location have two subclasses : (a) *measures of central values* such as mean, median and mode which give some central scores of the frequency

distribution, around which other individual scores of the sample tend to collect, and (b) *fractiles* such as percentiles and quartiles which give some points of the distribution below which lie specific fractions, or percentages of the total number of scores.

(ii) *Statistics of dispersion* : These are measures of the spread or scatter of the scores of a sample around a central value like mean or median. They are further classified as follows : (a) *absolute measures of dispersion* such as standard deviation, variance and quartile deviation, computed directly from the raw scores of the sample and expressed in either the same unit as the scores, or some power that unit, and (b) *relative measures of dispersion* such as coefficient of variation and coefficient of quartile deviation, which consist of expressions of some absolute measures of dispersion as ratios or percentages of the corresponding statistics of location, and do not bear the units of the raw scores.

(iii) *Statistics of correlation* : These measure and describe the magnitude and direction of relationship between two or more variables in the individuals of a sample, e.g., between body height and weight, wing length and trunk length, cardiac output and venous return, anxiety test scores and

performance, and intelligence test scores and achievement scores. Correlation coefficients such as product-moment correlation coefficient and rank-difference correlation coefficient belong to this class.

(b) *Sampling or inferential statistics* :

These include statistics like standard errors, which are not restricted within the limits of a sample unlike the descriptive statistics, rather go beyond the sample and help to make inferences and generalize them from the sample to the entire population. They find applications in testing of hypotheses, finding the significance of difference between statistics of different samples, and working out confidence intervals of parameters.

(c) *Prediction statistics* :

These include statistics such as regression coefficients and beta coefficients, used in predicting the most likely score of a dependent variable (criterion) in an individual from his actual score(s) in one or more independent variables (predictors). Examples : predictions of workshop skill scores from mechanical aptitude test scores, plasma prothrombin concentration from blood clotting time, body surface area from body height and weight, and O₂ consumption from tracheal ventilation.

GLOSSARY

- parameter** : any summary value or numerical index such as mean, median, standard deviation, standard error and correlation coefficient of all the scores of a variable or variables in the entire population.
- population** : the entire aggregate of all such existing individuals, cases or events in the universe as either possess a given form or belong to a given class of the specific variable investigated in a test, experiment or survey.
- sample** : a relatively small group of individuals, cases or events, chosen from a population for being subjected to an experiment, test or survey instead of that population, as its representative.
- sampling** : drawing of a sample by choosing individuals or cases from a population.
- sampling, judgement** : sampling by the deliberate and arbitrary choice of some individuals or cases from the population, depending on the personal judgement of the investigator.
- sampling, multistage** : simple random sampling separately at each of several stages formed by successive stepwise divisions of a vast widespread population.

- sampling, probability** : random choice of individuals or cases for a sample from the population, depending on the laws of probability.
- sampling, purposive** : random probability sampling from only such a limited section of the population as is adjudged by the investigator to represent the entire population.
- sampling, simple random** : random probability sampling from the undivided whole of a small, finite and homogeneous population.
- sampling, stratified random** : simple random samplings separately from each of the strata or subpopulations of a large heterogeneous population to ensure the random inclusion of individuals from each stratum into the sample.
- sampling, systematic** : choice of individuals for a sample at randomly predetermined intervals along a pre-arranged serial order of all members of the population.
- sampling distribution** : distribution of values of a statistic of different samples around the corresponding parameter.
- sampling error** : difference between a statistic and the corresponding parameter.
- scale, interval** : scale for a measurement variable like temperature, having no true zero point, starting consequently from any convenient arbitrary zero, and making it mathematically impermissible to compute a ratio of any two scores of such a scale.
- scale, ratio** : scale for a measurement variable such as length, mass and volume, having a true zero point and enabling the computation of a ratio between any two scores of such a scale.
- statistic** : any summary value or numerical index such as mean, standard deviation, quartile, quartile deviation and standard error, worked out from the scores of a variable in a sample and used as an estimate of the corresponding population parameter.
- statistics** : (a) plural of 'statistic' ; (b) the science of application of mathematical principles in the collection, presentation and analysis of the data of an experiment, test or survey, for interpretation and inference.
- statistics, descriptive** : statistics such as mean, standard deviation and correlation coefficient which describe the properties of a sample with respect to the relevant variable(s).
- statistics, prediction** : statistics such as regression coefficients and beta coefficients that are used in predicting the most likely value of one variable from the known values of one or more other variables.
- statistics, sampling** : statistics such as standard errors which go beyond a single sample to estimate sampling errors and to make inferences.
- statistics of correlation** : correlation coefficients that describe the magnitude and direction of relationship between the changes of two or more variables in a sample.
- statistics of dispersion** : descriptive statistics such as standard deviation, variance and coefficient of variation, which measure the scatter or dispersion of individual scores of a sample around a central value like mean or median.
- statistics of location** : descriptive statistics such as mean, median and mode, which describe the location of a specific point of the frequency distribution of sample scores on the scale of the variable.
- variable** : any property, phenomenon or event that varies in quantity or quality either spatially, i.e., from case to case or place to place even at the same instant, or temporally, i.e., from time to time even at the same location or in the same individual.

- variable, classification :** such an independent variable of an experiment as is beyond the control of the investigator and may thus suffer from random changes during the experiment.
- variable, continuous :** a variable that can have values, not only in whole numbers of units, but also in fractional units between two whole units.
- variable, dependent :** the variable whose possible change on exposure to specific independent variable(s) is being studied in an experiment or test.
- variable, discontinuous :** a variable that can have only certain specific values — usually in whole numbers of units only — and cannot possess intermediate values such as values in fractional units.
- variable, independent :** variable(s) whose possible effects on a dependent variable are being explored in a given experiment or test.
- variable, measurement :** variable whose magnitude can be measured or counted and expressed in numerical values.
- variable, nominal :** a variable such as sex, blood group, race and caste, which cannot be measured, counted or expressed quantitatively in numerical values, nor can the individuals of a sample be ranked with respect to such a variable.
- variable, ordinal :** a variable such as ferocity, timidity, personality, attentiveness and leadership quality, that varies distinctly in magnitude from individual to individual in a sample, enabling them to be ranked accordingly, but cannot be measured quantitatively in the individuals.
- variable, relevant :** any variable, not intended to be used by the investigator for studying its effect on the dependent variable, but nevertheless affecting the latter if not effectively controlled by the investigator.
- variable, sequence-relevant :** relevant variable(s) arising from variations in the sequence or order of application of the doses or levels of independent variable(s) in an experiment.
- variable, situational relevant :** relevant variable(s) arising from experimental situations and environment, such as variations in pH, temperature, light intensity, noise, humidity, work space, supervision in workshop, etc.
- variable, subject-relevant :** relevant variable(s) arising from variations of individuals of the sample, such as their differences in age, sex, body weight, motivation and intelligence.
- variable, treatment :** independent variable(s), strictly controlled by the investigator and not suffering from random changes during the experiment.

2. PRESENTATION OF DATA

Mere inspection of numerical scores of untreated raw data of an experiment or survey elicits little meaningful information and fails to help in the interpretation of observations. For understanding and analyzing the data, and for comparison with other sets of data, the raw data are systematically arranged, and graphically represented.

2.1 FREQUENCY DISTRIBUTIONS

A *frequency distribution* shows the number (frequency) of individuals, cases, events or scores of a sample, arranged and tabulated in the classes into which the variable under investigation has been classified.

Qualitative frequency distribution

A qualitative frequency distribution shows the number of individuals or cases of a sample in different classes of a nominal variable such as blood group, sex, phenotype, eye color, fur color, race, religion and occupation (Table 2.1). A nominal variable cannot be measured quantitatively in numerical units ; so, it is divided into classes depending only on the qualitative distinction between the individuals according to the presence or absence of a particular property or characteristic. The classes are, therefore, located only at specific points on the scale of the variable, separated from each other by intervening gaps ; no case falls in the gap between two such points. Such frequency distributions are thus known as *point distributions*. Classes of such a variable have no numerical range and are entered in one column of a *simple frequency table* while the counted frequency of individuals in each class is entered in another column against that particular class.

Table 2.1. A qualitative frequency distribution of blood groups in a sample.

Blood group	Frequency (f)
A	40
B	164
O	24
AB	26
Total (n)	254

Quantitative frequency distribution

This consists of a table showing the number (frequency) of individuals or cases of a sample in different classes or groups, into which the entire range of *scores* (numerical values) of a measurement variable has been divided. The data, arranged into such a frequency distribution, are called *grouped data* because the scores are distributed among the classified groups. Such classification reveals salient features of the variable in the sample in a meaningful way, e.g., the class having the highest frequency, and the pattern of distribution in different classes. It also helps in the statistical analysis and interpretation of data, particularly of a large sample.

1. Frequency distribution of a continuous variable :

There is no real gap between the classes of such a variable; the consecutive classes are continuous with each other.

(a) The *total range* of scores is worked out from the highest and lowest observed scores.

$$\text{Range} = (\text{highest score}) - (\text{lowest score}) + 1.$$

(b) To divide the total range into a suitable number of *class intervals*, the class size or range of scores of each class is so chosen that (i) each class interval covers about 3, 5 or 10 scores, the choice depending on the total range of scores and

the sample size, (ii) the class size should preferably be the same for all the class intervals, (iii) the total range is thus divided into 5-20 class intervals of equal size, (iv) the lowest and the highest scores of the data must fall within the lowest and the highest intervals respectively, and (v) the incidence of too many or too few cases in a class is avoided – details are obscured by too long class intervals with too many cases in some of them while the purpose of a meaningful arrangement of the data is defeated by too short class intervals with no or few cases in most of them.

(c) *Score limits* of each class, viz., its highest and lowest scores, should be stated precisely to avoid confusion in entering the observed scores in the respective intervals. In Table 2.3, for example, the sixth and seventh class intervals are given in an ascending order as 66-68 and 69-71 respectively. So, there should be no confusion about a score 69 belonging to the seventh interval only.

(d) In a continuous series of data, there should be no gap between either the limits of successive intervals or the consecutive scores. Each score is thus deemed to occupy a unit interval extending from 0.5 below that score to 0.5 above the latter. For example, the score 66 should be taken to extend from 65.5 to 66.5 and the score 67 from 66.5 to 67.5 so that no gap is left between 66 and 67. Therefore, *true class limits* or *class boundaries* like 65.5-68.5 and 68.5-71.5 should be used instead of the score limits like 66-68 and 69-71, respectively (Table 2.3). Each class interval then is not separated by gaps from the neighbouring class intervals.

True lower limit (X_l) = $\frac{1}{2}$ [(lower score limit of the interval) + (upper score limit of the next lower interval)].

True upper limit (X_u) = $\frac{1}{2}$ [(upper score limit of the interval) + (lower score limit of the next higher interval)].

(e) The *class size* (i) of an interval is obtained

from the difference between either its true limits, or the upper (or lower) score limits of consecutive intervals. Thus,

Either, $i = X_u - X_l$;

or, $i = (\text{upper score limit of one interval}) - (\text{upper score limit of the next lower interval})$.

(e) Class intervals may be tabulated either with the lowest interval at the bottom and the highest one at the top, or with the lowest interval at the top and the highest one at the bottom.

(f) In such a *grouped frequency distribution*, all individual cases are assumed to possess the score identical with the *midpoint* (X_c) of that interval. Where X_u and X_l are the true upper and lower limits respectively of the interval :

either $X_c = \frac{1}{2}(X_l + X_u)$;

or, $X_c = \frac{1}{2} [(\text{lower score limit}) + (\text{upper score limit})]$

For example, for a class interval of 66-68 (65.5-68.5), either, $X_c = \frac{1}{2}(65.5 + 68.5) = 67$, or, $X_c = \frac{1}{2}(66 + 68) = 67$.

(g) After tabulating the limits and midpoints of each interval, each individual case is entered as a *tally* against the interval to which it belongs; every fifth case in a class is entered as a diagonal tally against the last four tallies in that class (Table 2.3). Individual scores, coinciding with true limits of intervals, are all entered either in the intervals with true lower limits identical with the respective scores, or in the intervals with true upper limits coinciding with the respective scores. After all scores of the sample have been so entered as tallies against the respective intervals, the total frequency f in each class interval is entered in the table. The frequencies of all the intervals are finally totalled to give the sample size.

(h) When a few scores are scattered over wide ranges at one or both ends of the distribution, they are often grouped together in a single class interval with an indeterminate lower or upper limit (*open class interval*) with a range such as '15 and below' or '80 and above' for respectively the lowest and the highest classes. No midpoint can be computed

for open class intervals. Such frequency distributions are called *incomplete distributions*.

2. Frequency distribution of a discrete or discontinuous variable :

Because consecutive scores of a discrete variable like RBC count and family size are separated by real gaps (page 3), the classes of such a variable are separated by intervening gaps. So, discrete frequency distribution has its class intervals with score limits in whole units only, and with no true class limit in fractional or decimal units. This leaves gaps between the classes with no score in these gaps.

Where scores of a discrete variable are relatively few, each occurring once or only a few times in the sample, their frequencies are arranged in an ungrouped frequency distribution called a *simple frequency table*. Each single distinct score is entered individually in this table as if to

constitute a class by itself ; no class interval is formed here by grouping a range of more than one score. The frequency of a particular score is entered directly against it and not in any grouped interval of a range of scores. In the case cited in Table 2.2, no family can exist with a fractional number of children like 2.5 or 3.3, real gaps exist between the whole numbers of children, and the frequencies are entered against individual scores, not against grouped intervals.

Table 2.2. A simple frequency table for a discrete frequency distribution of families in terms of the number of children.

Number of children	Number of families
0	7
1	35
2	67
3	43
5	10
Total	162 (n)

Example 2.1.1.

Tabulate the following scores of body weight (kg) in a sample in the form of a frequency distribution :

68, 59, 57, 64, 52, 60, 62, 57, 61, 61, 71, 51, 55, 54, 65, 67, 54, 62, 58, 60, 54, 62, 65, 71, 63, 60, 61, 56, 67, 64, 57, 61, 60, 62, 59, 57, 64, 58, 61, 63, 62, 62, 60, 58, 67, 63, 64, 57, 61, 60, 65, 67, 70, 58, 51, 61, 62, 65, 52, 60, 55, 63, 62, 60, 67, 55, 62, 61, 64, 59.

Solution :

Sample size (n) = 70.

Highest score = 71. Lowest score = 51. Range = $(71 - 51) + 1 = 21$.

Proposed size of interval (i) = 3. Number of classes (k) = $\frac{\text{range}}{\text{interval size}} = \frac{21}{3} = 7$.

Table 2.3. Tabulation of a continuous frequency distribution of body weight scores.

Class intervals		Midpoints	Tallies	Frequency
Score limits	True limits	(X_c)		(f)
51-53	50.5-53.5	52		4
54-56	53.5-56.5	55		7
57-59	56.5-59.5	58		12
60-62	59.5-62.5	61		25
63-65	62.5-65.5	64		13
66-68	65.5-68.5	67		6
69-71	68.5-71.5	70		3
Total				70 (n)

Body weight being a continuous variable, a continuous frequency distribution is tabulated in Table 2.3.

(a) The true lower and upper class limits (X_l and X_u) of each interval are calculated and entered. For example, for the interval 57-59,

$$X_l = \frac{1}{2} [(lower\ score\ limit\ of\ the\ interval) + (upper\ score\ limit\ of\ the\ next\ lower\ interval)] \\ = \frac{1}{2} (57 + 56) = 56.5.$$

$$X_u = \frac{1}{2} [(upper\ score\ limit\ of\ the\ interval) + (lower\ score\ limit\ of\ the\ next\ higher\ interval)] \\ = \frac{1}{2} (59 + 60) = 59.5.$$

(b) The midpoint X_c of each interval is computed by the following formula :

$$X_c = \frac{1}{2} [(lower\ score\ limit) + (upper\ score\ limit)]$$

For example, for the interval 57-59,

$$X_c = \frac{1}{2}(57 + 59) = 58.$$

(c) Each individual score is entered as a tally against its class interval, each fifth score of a class being entered as a diagonal tally against the preceding four tallies in that interval. The total number of tallies in each class is then entered as the frequency (f) of that class.

Example 2.1.2.

Tabulate the following femur length scores ($mm \times 10^{-2}$) of a sample of aphids in a frequency distribution :

31, 36, 37, 32, 43, 40, 34, 42, 36, 35, 39, 44, 37, 42, 44, 35, 45, 40, 31, 33,
43, 34, 38, 39, 37, 35, 37, 36, 42, 49, 46, 38, 36, 31, 48, 37, 36, 37, 40, 38.

Solution :

Sample size (n) = 40.

Highest score = 49. Lowest score = 31. Range = $(49 - 31) + 1 = 19$.

Proposed interval size (i) = 3. Number of classes = $\frac{range}{interval\ size} = \frac{19}{3} = 6.33 \approx 7$.

Table 2.4. Tabulation of a continuous frequency distribution of aphid femur length scores.

Class intervals		Midpoints (X_c)	Tallies	Frequency (f)
Score limits	True limits			
31-33	30.5-33.5	32		5
34-36	33.5-36.5	35		10
37-39	36.5-39.5	38		11
40-42	39.5-42.5	41		6
43-45	42.5-45.5	44		5
46-48	45.5-48.5	47		2
49-51	48.5-51.5	50		1
Total				40 (n)

Femur length being a continuous variable, the scores are arranged in a continuous frequency distribution (Table 2.4).

(a) The true lower and upper class limits (X_l and X_u) of each interval are worked out and entered in Table 2.4.

For example, for the interval 40-42,

$$X_l = \frac{1}{2}[(\text{lower score limit of the interval}) + (\text{upper score limit of the next lower interval})] \\ = \frac{1}{2}(40 + 39) = 39.5.$$

$$X_u = \frac{1}{2}[(\text{upper score limit of the interval}) + (\text{lower score limit of the next higher interval})] \\ = \frac{1}{2}(42 + 43) = 42.5.$$

(b) The midpoint (X_c) of each interval is worked out and entered. For example, for the interval 40-42,

$$X_c = \frac{1}{2}[(\text{lower score limit}) + (\text{upper score limit})] \\ = \frac{1}{2}(40 + 42) = 41.$$

(c) Each individual score is entered as a tally against its class interval, each fifth score of a class being entered as a diagonal tally across the preceding four tallies in that interval. The total number of tallies in each class is entered as the frequency (f) in that class.

Example 2.1.3.

Tabulate the following respiratory rate scores (per minute) of a sample of orang-utans in a frequency distribution:

21, 16, 8, 12, 14, 10, 17, 17, 20, 13, 15, 19, 25, 16, 15, 23, 14, 18, 16, 17.

Solution:

Sample size (n) = 20.

Highest score = 25. Lowest score = 8. Range = $(25 - 8) + 1 = 18$.

Proposed interval size (i) = 4. Number of classes = $\frac{\text{range}}{\text{interval size}} = \frac{18}{4} = 4.5 \approx 5$.

Table 2.5. Tabulation of a discontinuous frequency distribution of orang-utan respiratory rate scores.

Class intervals (Score limits)	Midpoints (X_c)	Tallies	Frequency (f)
8-11	9.5	II	2
12-15	13.5	III I	6
16-19	17.5	III III	8
20-23	21.5	III	3
24-27	25.5	I	1
Total			20 (n)

Respiratory rate being a discontinuous variable, the scores are arranged in a discrete frequency distribution (Table 2.5).

(a) No true limit is worked out for any class interval because the variable is a discrete one.

(b) The midpoint (X_c) of each interval is worked out and entered. For example, for the interval 16-19,

$$X_c = \frac{1}{2}[(\text{lower score limit}) + (\text{upper score limit})] \\ = \frac{1}{2}(16 + 19) = 17.5.$$

(c) Each individual score is entered as a tally against its class interval, each fifth score of a class being entered as a diagonal tally across the preceding four tallies in that interval. The total number of tallies in each class is entered as the frequency (f) in that class.

2.2 PIE DIAGRAM

Pie diagrams are used for the graphical representation of frequency distributions of nominal variables. The diagram consists of a circle (pie) with its entire area representing the sample size or total frequency (n) of cases. The angle 360° at its centre is divided by a protractor to cut the circle into segments, each proportional in area to the relative frequency of cases in a class of the variable. The angle θ° to be cut off from 360° for a segment is given by : $\theta = 360^\circ \times f/n$, where f is

the frequency of cases in the relevant class and n is the sample size. The areas of different segments thus represent graphically the proportions of the net frequency (n) in the respective classes. But pie diagrams are inconvenient for comparing more than one sample because (i) separate pie diagrams have to be used for different samples, and (ii) if there are either too many classes or too low frequencies in some classes, the segments for such classes are too narrow for precise drawing and correct comparison.

Example 2.2.1.

Draw a pie diagram for the following frequency distribution of blood groups in a sample.

Blood groups	:	O	A	B	AB	Total
Frequencies	:	258	172	387	43	860

Solution :

The entire 360° of the pie represents the total frequency (n) of 860. Where θ° is the angle of a segment and f is the frequency of cases in the class represented by that segment,

$$\theta = 360^\circ \times \frac{f}{n} ;$$

$$\therefore \text{for O group, } \theta = 360^\circ \times \frac{258}{860} = 108^\circ ; \quad \text{for A group, } \theta = 360^\circ \times \frac{172}{860} = 72^\circ ;$$

$$\text{for B group, } \theta = 360^\circ \times \frac{387}{860} = 162^\circ ; \quad \text{for AB group, } \theta = 360^\circ \times \frac{43}{860} = 18^\circ .$$

Using a protractor, segments of 108° , 72° , 162° and 18° are cut off successively from the circle to represent the frequencies of O, A, B and AB groups, respectively (Fig. 2.1). The segments are shaded or colored distinctly from each other and labelled.

2.3 BAR DIAGRAM

A bar diagram consists of set(s) of bars or columns, used for the graphic representation and comparison of the classwise frequency distribution(s) of a nominal or discrete variable in one or more samples.

Simple bar diagram

This consists of a set of several parallel bars or rectangles, one for each group or class of the

variable. The bars may be *vertical* or *horizontal*, have *equal widths* chosen arbitrarily by the investigator, and are separated from each other by small *intervening gaps* indicating that real gaps exist between the classes of the variable. Frequencies, amounts, percentages, etc., are scaled parallel to the lengths of the bars, starting from a zero value to avoid any misleading impression about the relative lengths of the bars. The length or height of each bar is made to correspond to the

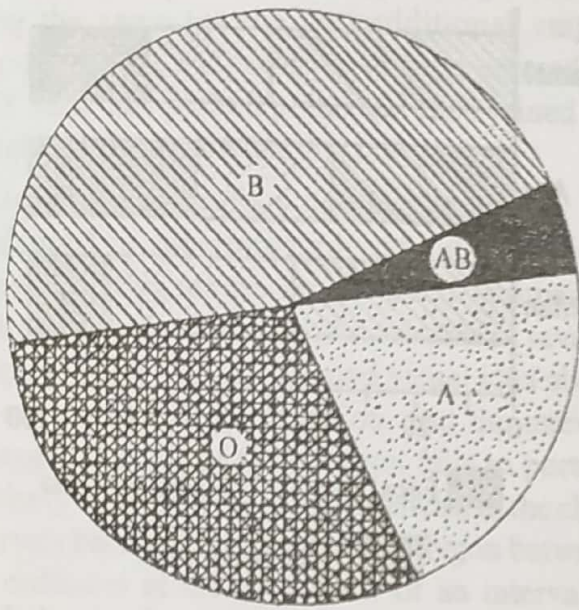


Fig. 2.1. A pie diagram of frequency distribution of blood groups in a sample.

frequency, amount or percentage in the relevant class. Because the bars are of equal widths, their areas are directly proportional to their respective lengths and consequently to the frequencies, amounts or percentages in the respective classes. The distribution in different classes may be studied by comparing the lengths (or areas) of their respective bars (Fig. 2.2).

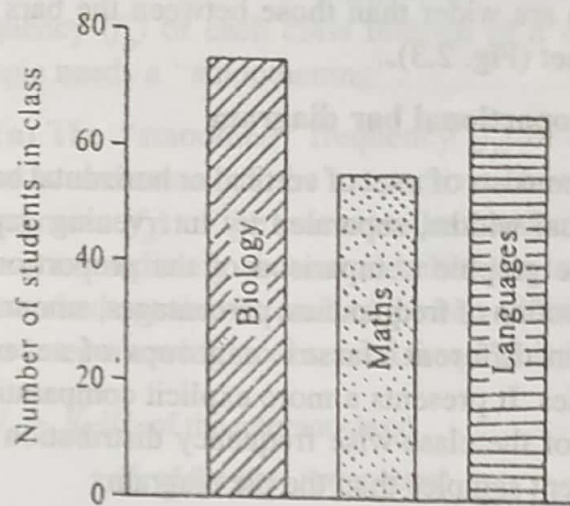


Fig. 2.2. A simple bar diagram showing the frequency distribution of students of different subjects in a college.

Multiple bar diagram

To show the frequency distributions in the groups of more than one sample, a *multiple bar diagram* is drawn with as many sets of bars as the number of samples. The bars of each set show the frequency distribution of a particular sample in the classes of the variable and are as many as the classes. The set of bars of each sample is separated from those of neighbouring samples by gaps.

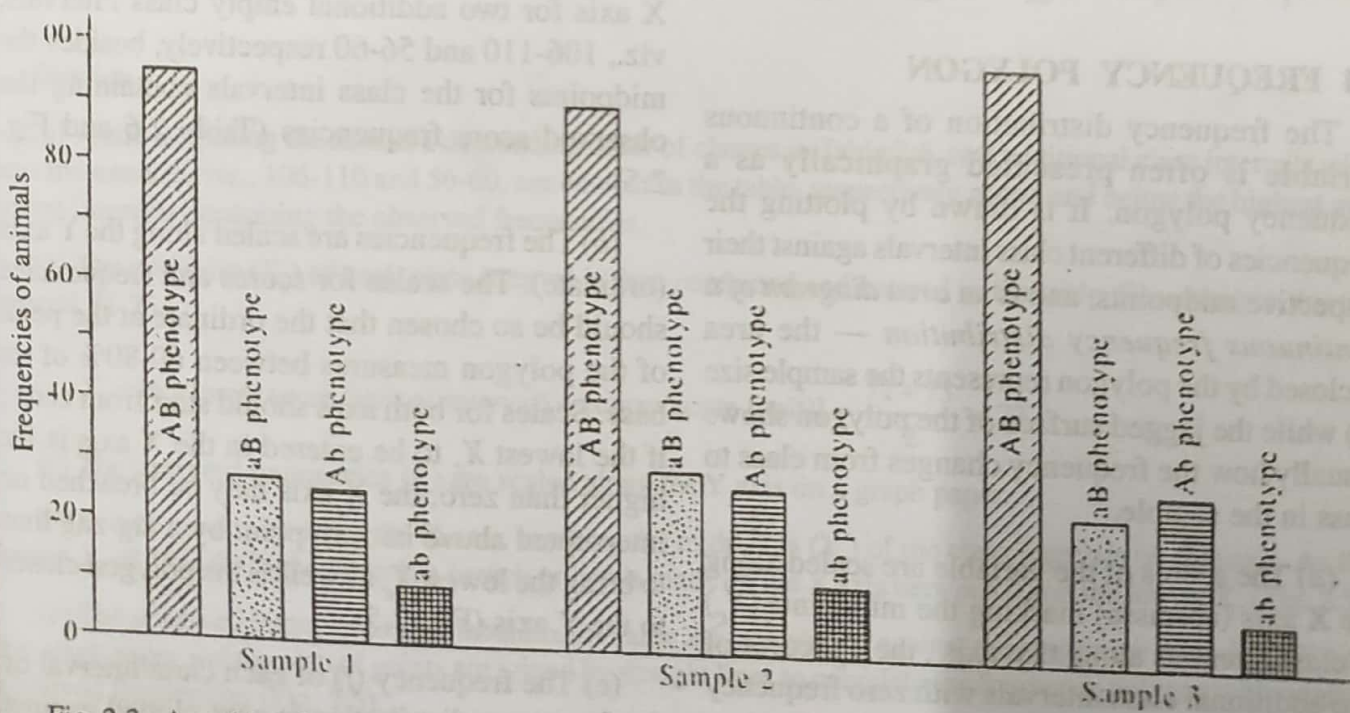


Fig. 2.3. A multiple bar diagram showing the distributions of *Drosophila* phenotypes in three samples.

which are wider than those between the bars of each set (Fig. 2.3).

Proportional bar diagram

It consists of a set of vertical or horizontal bars of equal widths, separated by intervening gaps, for the graphic comparison of the proportional distribution of frequencies, percentages, amounts, etc., in different classes or groups of several samples. It presents a more explicit comparative view of the class-wise frequency distribution in different samples than the pie diagram.

Each sample is assigned a bar, the entire area of which represents the total frequency (n) of that sample. The frequencies, percentages, amounts, etc., are scaled along an axis parallel to the lengths of the bars (Fig. 2.4). Each bar is divided into as many segments as the number of classes or groups. The lengths of the segments are determined by the frequencies, percentages or amounts in the relevant classes or groups. Consequently, their areas are proportional to the frequencies, percentages or amounts in the respective classes or groups. For easy comparison, the classes or groups are arranged in the same order and shaded or colored similarly in all the bars.

2.4 FREQUENCY POLYGON

The frequency distribution of a continuous variable is often presented graphically as a frequency polygon. It is drawn by plotting the frequencies of different class intervals against their respective midpoints, and is an *area diagram of a continuous frequency distribution* — the area enclosed by the polygon represents the sample size (n) while the jagged surface of the polygon shows visually how the frequency changes from class to class in the sample.

(a) The scores of the variable are scaled along the X axis (abscissa) marking the midpoints (X_c) of class intervals along that axis; the X_c scores of two additional class intervals with zero frequency are also entered on the X axis, one for the class

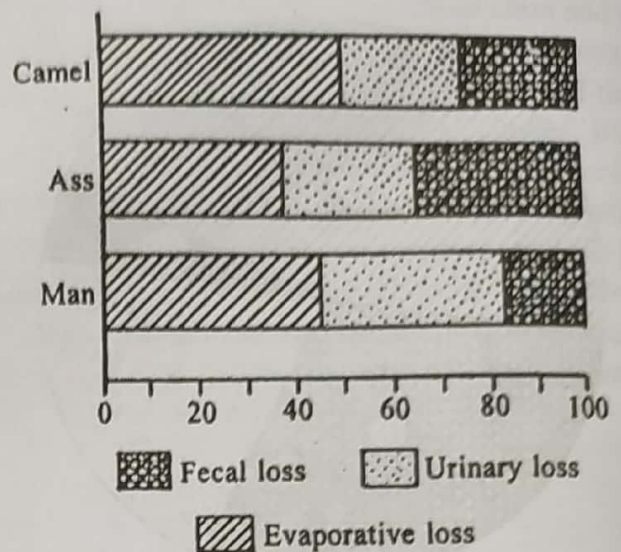


Fig. 2.4. A proportional bar diagram showing the loss of body water through different channels in three mammalian species as percentages of their total daily water loss.

just below the lowest interval containing observed frequencies, and the other for the class interval just above the highest interval with observed frequencies — this would enable the polygon to reach the base line or zero frequency level at both ends. For instance, in *Example 2.4.1*, the midpoints 108 and 58 have been included in the X axis for two additional empty class intervals, viz., 106-110 and 56-60 respectively, besides the midpoints for the class intervals containing the observed score frequencies (Table 2.6 and Fig. 2.5).

(b) The frequencies are scaled along the Y axis (ordinate). The scales for scores and frequencies should be so chosen that the ordinate at the peak of the polygon measures between 60-80% of its base. Scales for both axes should start from zero; if the lowest X_c to be entered in the X axis is far higher than zero, the X axis may be breached or interrupted above its zero point by a zig-zag line to bring the lowest X_c as well as the polygon closer to the Y axis (Fig. 2.5).

(c) The frequency (f) of each class interval of the frequency distribution is next plotted against

the midpoint (X_c) of the corresponding interval, doing the same for the two additional empty intervals whose midpoints have been entered in the X axis. The plotted points are then joined by straight lines to complete the polygon.

Because more than one polygon may be overlapped or superimposed without confusion, frequency polygons are very useful in comparing the frequency distributions of a variable in several samples. They also give a good visual idea about the contours of the distribution. But the uneven or jagged surface of the polygon fails to portray precisely the proportionate frequencies in the class intervals because the area of the polygon between the ordinates at the two limits of an interval is hardly proportional to the frequency in the latter.

Smoothed frequency polygon :

The smaller the sample, the more is the jaggedness of the polygon. To make the polygon

for a small sample less jagged, the observed frequency (f_o) of each class interval of a small sample needs a "smoothing".

(a) The "smoothed" frequency (f_s) of each class is the mean of the respective observed frequencies (f_o) of the relevant class and of the classes immediately above and below the latter ; f_s is worked out also for each of the two additional class intervals with no observed frequencies.

$$f_s = \frac{1}{3} [(f_o \text{ of the relevant class}) \\ + (f_o \text{ of the next lower class}) \\ + (f_o \text{ of the next higher class})]$$

(b) The computed smoothed frequencies are plotted against the midpoints of the respective class intervals. The plotted points are joined by straight lines to give the smoothed frequency polygon which, however, would not reach the X axis at its two ends (Fig. 2.5).

Example 2.4.1.

Draw the observed and smoothed frequency polygons of diastolic blood pressure (mm Hg) using the following diastolic BP scores in a sample.

Class intervals	:	61-65	66-70	71-75	76-80	81-85	86-90	91-95	96-100	101-105
Frequencies	:	2	7	12	23	40	22	15	8	1

Solution :

(a) While tabulating the data in a descending order of classes in Table 2.6, two additional class intervals with zero frequencies, viz., 106-110 and 56-60, are entered in the table, respectively above and below the highest and lowest intervals containing the observed frequencies.

(b) The midpoint (X_c) of each class interval is then computed and entered in the table. For example, for the interval 71-75,

$$X_c = \frac{1}{2} [(lower \text{ score limit of interval}) + (upper \text{ score limit})] = \frac{71 + 75}{2} = 73.$$

(c) The observed frequencies (f_o) are scaled along the Y axis on a graph paper.

(d) Scores are scaled along the X axis, marking the midpoints (X_c) of the class intervals on that axis. As the lowest X_c of 58 is far higher than 0, a breach or gap is made on the X axis between 0 and the lowest X_c (Fig. 2.5).

(e) The observed frequencies (f_o), tabulated in Table 2.6, are plotted against the respective midpoints (X_c) on the graph paper and the plotted points are joined by straight lines to complete the frequency polygon with original observed frequencies (Fig. 2.5).

Table 2.6. Distributions of observed and smoothed frequencies of diastolic BP.

Class intervals	X_c	f_0	f_s
106-110	108	0	$\frac{1}{3}(0+1+0) = 0.3$
101-105	103	1	$\frac{1}{3}(1+8+0) = 3.0$
96-100	98	8	$\frac{1}{3}(8+15+1) = 8.0$
91-95	93	15	$\frac{1}{3}(15+22+8) = 15.0$
86-90	88	22	$\frac{1}{3}(22+40+15) = 25.7$
81-85	83	40	$\frac{1}{3}(40+23+22) = 28.3$
76-80	78	23	$\frac{1}{3}(23+12+40) = 25.0$
71-75	73	12	$\frac{1}{3}(12+7+23) = 14.0$
66-70	68	7	$\frac{1}{3}(7+2+12) = 7.0$
61-65	63	2	$\frac{1}{3}(2+0+7) = 3.0$
56-60	58	0	$\frac{1}{3}(0+0+2) = 0.7$

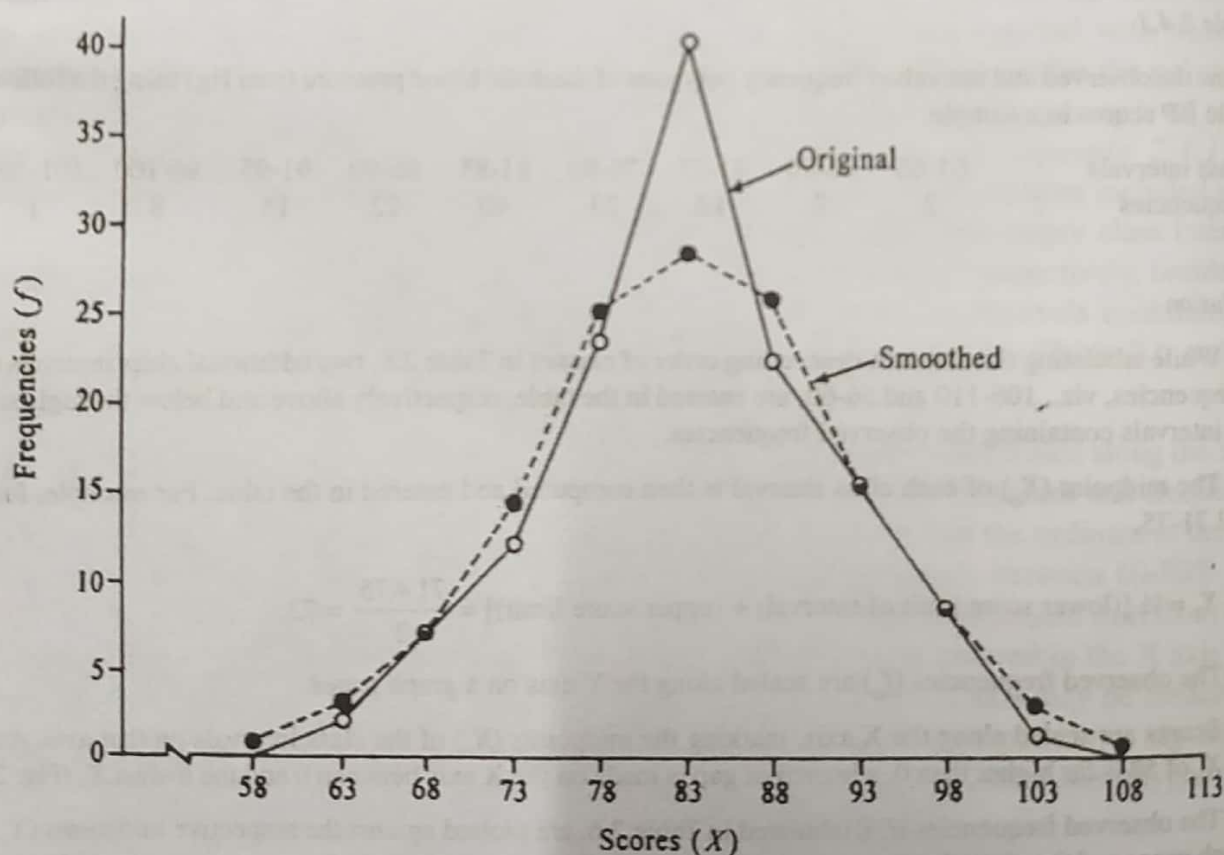


Fig. 2.5. Original and smoothed frequency polygons of the distribution of diastolic blood pressure in a sample.

(f) The smoothed frequency (f_s) is worked out for each class interval and entered in Table 2.6. For example, for the class interval 101-105,

$$\begin{aligned} f_s &= \frac{1}{3} [(f_o \text{ of the relevant class}) + (f_o \text{ of the next lower class}) + (f_o \text{ of the next higher class})] \\ &= \frac{1}{3}(1 + 8 + 0) = 3.0. \end{aligned}$$

(g) Each f_s is plotted against the corresponding X_c on a graph paper and the plotted points are joined by straight lines to give the smoothed frequency polygon (Fig. 2.5).

Example 2.4.2.

Draw the observed and smooth frequency polygons of femur lengths ($\text{mm} \times 10^{-2}$) using the following femur length data of a sample of aphids.

Class intervals	:	31-33	34-36	37-39	40-42	43-45
Frequencies	:	5	12	20	9	4

Solution :

(a) While tabulating the data in a descending order of classes in Table 2.7, two additional class intervals with zero frequencies, viz., 46-48 and 28-30, are entered in the table, respectively above and below the highest and lowest intervals containing observed frequencies.

Table 2.7. Distributions of observed and smoothed frequencies of aphid femur lengths.

Class intervals	X_c	f_o	f_s
46-48	47	0	$\frac{1}{3}(0 + 4 + 0) = 1.3$
43-45	44	4	$\frac{1}{3}(4 + 9 + 0) = 4.3$
40-42	41	9	$\frac{1}{3}(9 + 20 + 4) = 11.0$
37-39	38	20	$\frac{1}{3}(20 + 12 + 9) = 13.7$
34-36	35	12	$\frac{1}{3}(12 + 5 + 20) = 12.3$
31-33	32	5	$\frac{1}{3}(5 + 0 + 12) = 5.7$
28-30	29	0	$\frac{1}{3}(0 + 0 + 5) = 1.7$

(b) The midpoint (X_c) of each class interval is worked out and entered in the table. For example for the interval 37-39,

$$X_c = \frac{1}{2} [(\text{lower score limit of interval}) + (\text{upper score limit})] = \frac{37 + 39}{2} = 38.$$

(c) The observed frequencies (f_o) are scaled along the Y axis on a graph paper.

(d) Scores are scaled along the X axis, marking the midpoints (X_c) of the class intervals on that axis. As the lowest X_c of 29 is far higher than 0, a breach or gap is made on the X axis between 0 and the lowest X_c (Fig. 2.6).

(e) The f_o scores of Table 2.7 are plotted against the respective X_c scores on the graph paper and the plotted points are joined by straight lines to complete the polygon with original observed frequencies (Fig. 2.6).

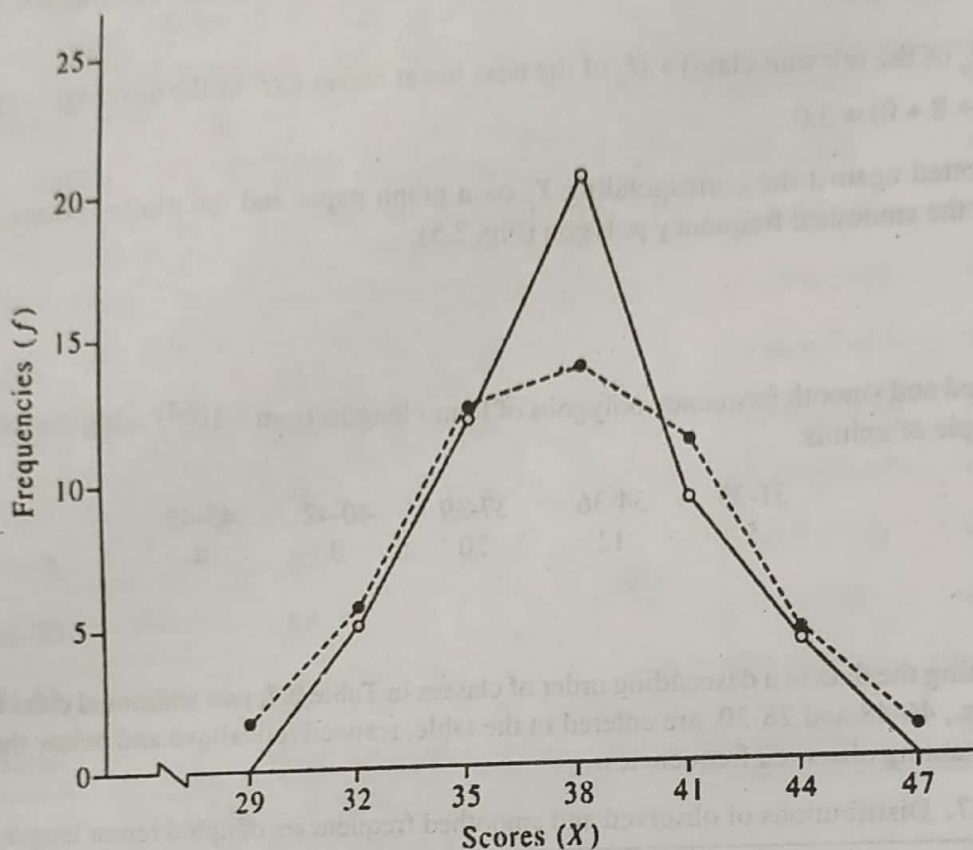


Fig. 2.6. Original and smoothed frequency polygons of aphid femur length distribution.

(f) The smooth frequency (f_s) is worked out for each class interval and entered in Table 2.7. For example, for the class interval 37-39,

$$f_s = \frac{1}{3} [(f_o \text{ of the relevant class}) + (f_o \text{ of the next lower class}) + (f_o \text{ of the next higher class})]$$

$$= \frac{1}{3} (20 + 12 + 9) = 13.7.$$

(g) Each f_s score is plotted against the corresponding X_c on a graph paper and the plotted points are joined by straight lines to give the smoothed frequency polygon (Fig. 2.6).

2.5 HISTOGRAM

Histogram or column diagram is a graphic representation of the frequency distribution of a continuous variable. It consists of a continuous set of bars, not separated by intervening spaces and thus indicating the absence of any real gap in the scale of the relevant variable. It is an *area diagram of a continuous frequency distribution*; its total area represents the sample size (n) while the area of each bar is proportional to the frequency of cases in a particular class interval. Unlike the frequency polygon where all the cases

in a class interval are shown to be located at its midpoint, the frequency of each class is shown here as uniformly distributed over the entire interval length. The consequent absence of the jagged appearance makes it more convenient than the frequency polygon for a visual portrayal of the proportionate frequencies in the class intervals. The outline of upper surfaces of the bars gives also a visual idea of the shape of the frequency curve. But the histogram is less convenient than the polygon in comparing the frequency distributions of more than one sample because as

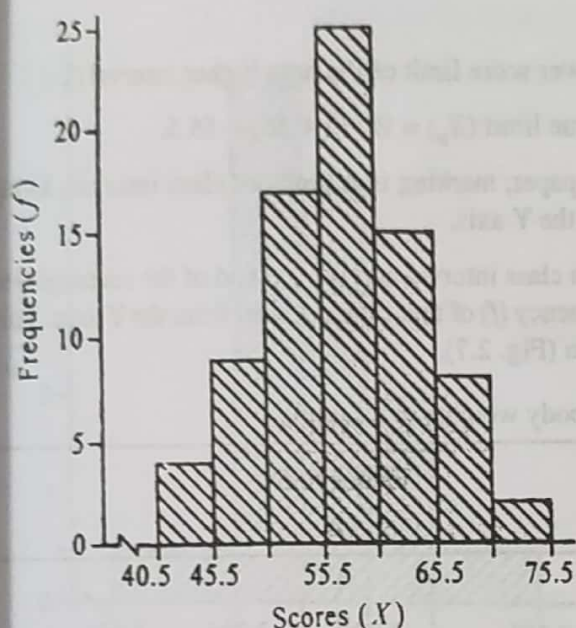


Fig. 2.7. Histogram of a frequency distribution of body weights.

many separate histograms have to be used as the number of samples — they cannot be superimposed on each other.

(a) To avoid intervening gaps between consecutive class intervals, *true class limits*, computed from the respective score limits of all the classes, have to be used in plotting the histogram. But *no additional class interval with zero frequency* is included beyond the intervals having observed frequencies.

(b) The scores (X) of the variable are scaled along the X axis on a graph paper and the true class limits of all the intervals of the frequency distribution are marked on the X axis. If the lower true limit of the lowest interval is far higher than

0, the X axis may be breached or interrupted by a zig-zag line between 0 and that true limit — this would bring the histogram closer to the Y axis (Fig. 2.7).

(c) The frequencies (f) are scaled along the Y axis. The scales for X and f should be so chosen that the ordinate for the highest f, i.e., the height of the tallest bar, measures between 60-80% of the base of the histogram.

(d) Two ordinates are raised on the X axis at the true limits of each class interval, and the top end of the rectangle being formed is closed by a horizontal line at the level of the frequency (f) of that class interval indicated by the Y scale. A bar is thus formed with its base extending over the length of the class interval and its height corresponding to the frequency of cases in that interval. This is repeated for all the class intervals in the data to draw a set of bars. As long as the class intervals and so, the bases of the bars are of equal lengths, the areas of the bars are proportionate to their heights and so, to the frequencies in the respective intervals.

For *unequal-size class intervals*, the bars differ in the width of their bases and their areas fail to indicate the proportional frequencies in the intervals. To remedy this, the frequency (f) of each interval is divided by its class size (i) to give its frequency density (f/i), i.e., the average frequency for unit length. Each bar is then drawn with its height equalling its f/i and its base coinciding with the original class size (i) of that interval. The areas of such bars are proportionate to the frequencies in the respective unequal intervals (Fig. 2.8).

Example 2.5.1.

Draw the histogram of the following frequency distribution of body weights (kg) in a sample :

Class intervals :	41-45	46-50	51-55	56-60	61-65	66-70	71-75
Frequencies :	4	9	17	25	15	8	2

Solution :

(a) The data are tabulated in Table 2.8; the true limits (X_l and X_u) of the class intervals are computed and

entered. For example, for the interval 51-55,

true limit = $\frac{1}{2}$ [(upper score limit of an interval) + (lower score limit of the next higher interval)].

lower true limit (X_l) = $\frac{1}{2}$ (50 + 51) = 50.5 ; upper true limit (X_u) = $\frac{1}{2}$ (55 + 56) = 55.5.

(b) The scores (X) are scaled along the X axis on a graph paper, marking true limits of class intervals along that axis (Fig. 2.7), while the frequencies (f) are scaled along the Y axis.

(c) Ordinates are raised on the X axis at true limits of each class interval and the top end of the rectangle, to be formed, is closed by a horizontal line at the level of the frequency (f) of that interval, read from the Y axis. This is repeated for all the class intervals to complete the histogram (Fig. 2.7).

Table 2.8. Frequency distribution of body weights in a sample.

Class intervals		Frequencies
Score limits	True limits	(f)
41-45	40.5-45.5	4
46-50	45.5-50.5	9
51-55	50.5-55.5	17
56-60	55.5-60.5	25
61-65	60.5-65.5	15
66-70	65.5-70.5	8
71-75	70.5-75.5	2
Total		80 (n)

Example 2.5.2.

Draw a histogram of the following frequency distribution of body heights (cm) in a sample :

Class intervals :	151-160	161-165	166-170	171-175	176-183
Frequencies :	5	25	20	10	4

Solution :

The class intervals of this frequency distribution are of unequal lengths.

(a) The class intervals and observed frequencies (f) are tabulated in Table 2.9 and the true limits of the class intervals are computed and entered as in Example 2.5.1.

Table 2.9. Distribution of frequency densities in unequal class intervals.

Class intervals		Size	f	f/i
Score limits	True limits	of interval (i)		
151-160	150.5-160.5	10	5	0.5
161-165	160.5-165.5	5	25	5
166-170	165.5-170.5	5	20	4
171-175	170.5-175.5	5	10	2
176-183	175.5-183.5	8	4	0.5

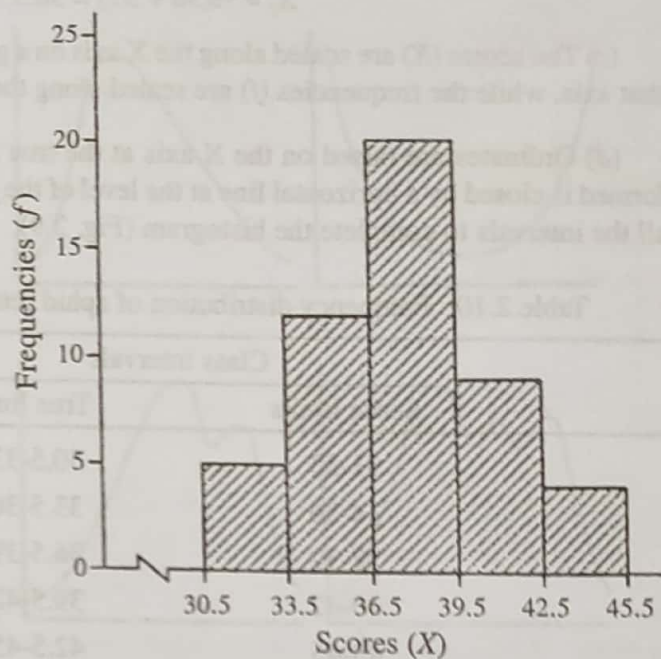
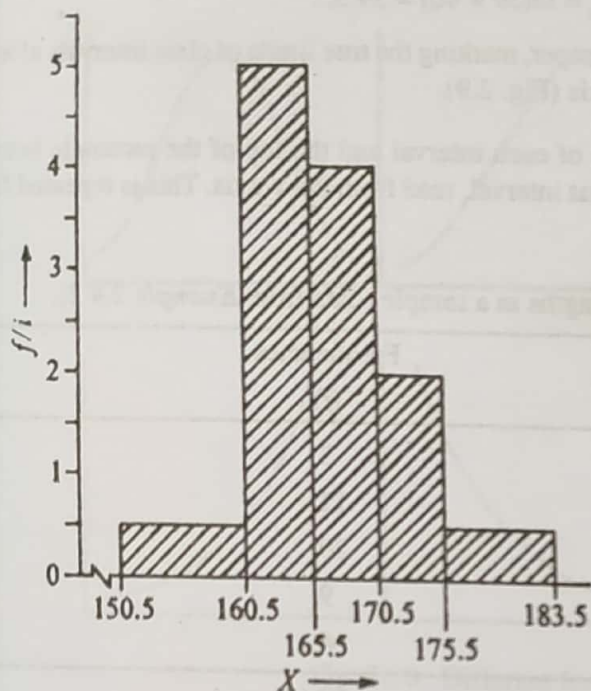


Fig. 2.8. Histogram of a frequency distribution of heights with unequal class intervals.

Fig. 2.9. Histogram of a frequency distribution of aphid femur lengths.

(b) The frequency f of each class interval is divided by the size i of that interval to get the frequency density f/i of the interval. For example, for the interval 151-160,

$$i = 10 ; f = 5 ; \therefore f/i = 5/10 = 0.5.$$

These frequency densities (f/i) are entered in Table 2.9 against the respective class intervals.

(c) The scores (X) are scaled along the X axis on a graph paper, marking the true limits of class intervals along that axis (Fig. 2.8).

(d) The frequency densities (f/i) are scaled along the Y axis.

(e) Ordinates are raised on the X axis at the true limits of each interval and the top of the rectangle being formed is closed by a horizontal line at the level of the f/i of that interval, read from the Y axis. This is repeated for all the class intervals to complete the histogram (Fig. 2.8).

Example 2.5.3.

Draw a histogram of the aphid femur length scores ($\text{mm} \times 10^{-2}$) of the data given in Example 2.4.2.

Solution :

(a) The data are entered in Table 2.10 ; no additional class interval with zero frequency is entered beyond the class intervals having f_o scores of the given data.

(b) The true limits (X_l and X_u) of each class interval are computed and entered in Table 2.10. For example, for the class interval 37-39,

$$\text{true limit} = \frac{1}{2} [(\text{upper score limit of an interval}) + (\text{lower score limit of the next higher interval})] ;$$

$$X_l = \frac{1}{2}(36 + 37) = 36.5; \quad X_u = \frac{1}{2}(39 + 40) = 39.5.$$

(c) The scores (X) are scaled along the X axis on a graph paper, marking the true limits of class intervals along that axis, while the frequencies (f) are scaled along the Y axis (Fig. 2.9).

(d) Ordinates are raised on the X axis at the true limits of each interval and the top of the rectangle being formed is closed by a horizontal line at the level of the f of that interval, read from the Y axis. This is repeated for all the intervals to complete the histogram (Fig. 2.9).

Table 2.10. Frequency distribution of aphid femur lengths in a sample (data from Example 2.4.2).

Class intervals		Frequencies
Score limits	True limits	(f)
31-33	30.5-33.5	5
34-36	33.5-36.5	12
37-39	36.5-39.5	20
40-42	39.5-42.5	9
43-45	42.5-45.5	4
Total		50 (n)

2.6 FREQUENCY DISTRIBUTION CURVE

With a very large sample, very short and equal class intervals, and scores expressed upto very fine fractions of a unit, the steplike or jagged outline of a histogram or a frequency polygon changes into a smooth curve. The graph of the frequency distribution thus becomes a *frequency distribution curve*, with its scores (X) and frequencies (f) scaled along the abscissa and the ordinate, respectively. The total area under the curve represents the sample size (n). The curve may be *unimodal*, *bimodal* or *multimodal* according as it has one, two or many peaks (Fig. 2.10). It may be either bilaterally symmetrical or asymmetric (*skewed*) with one tail more drawn out than the other. It may be bell-shaped, J-shaped, U-shaped or reverse J-shaped. All these depend on the pattern of frequency distribution in the sample or the population.

2.7 OGIVES

Ogives are graphical forms of distributions of cumulative frequencies (cf) or cumulative percentages (cP) of the scores of a sample.

Cumulative frequency

The *cumulative frequency* is the sum of the frequencies of all the scores of a continuous variable in a sample, either upto a particular score (*less-than cf*) or above a particular score (*more-than cf*). However, the term cumulative frequency or *cf* is generally used to mean the less-than type only, unless preceded by the word "more-than".

In grouped data, the *cf* (less-than type) of the k th class interval is the sum of the frequencies (f) of all the k number of intervals from the lowest one to the upper limit of the k th interval under consideration. Where cf_1, cf_2, \dots, cf_k are the cumulative frequencies and f_1, f_2, \dots, f_k are the observed frequencies of the successive class intervals in an ascending order in a frequency distribution,

$$cf_1 = f_1; \quad cf_2 = f_1 + f_2; \quad cf_3 = f_1 + f_2 + f_3; \\ cf_k = f_1 + f_2 + f_3 + \dots + f_k = n.$$

Cumulative frequencies are used in the computation of percentiles, percentile ranks, deciles, quartiles and median of a sample.

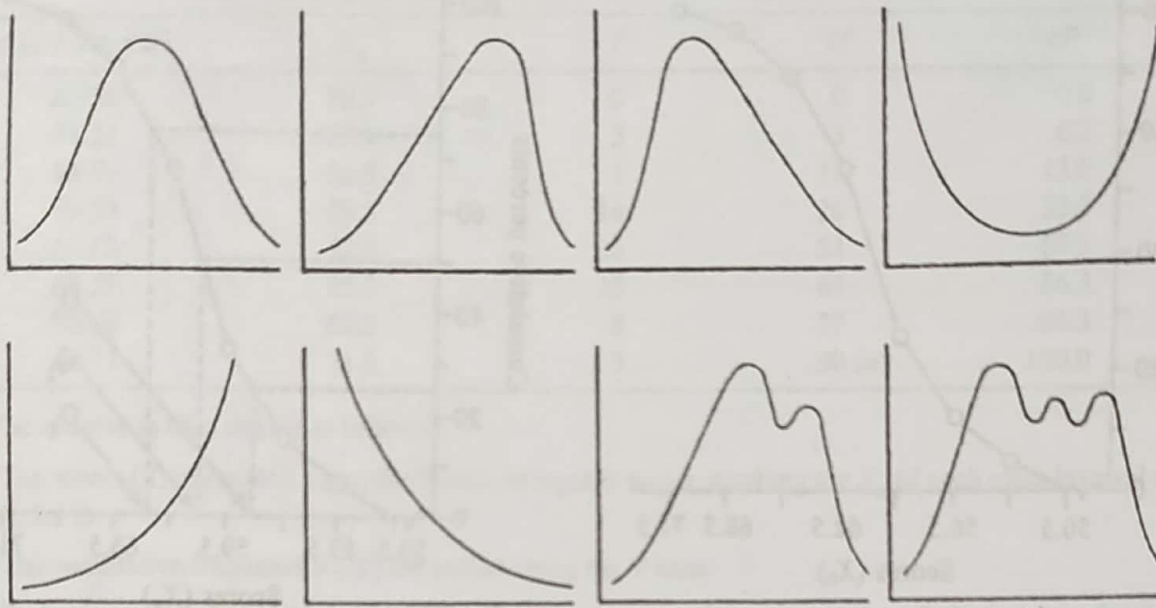


Fig. 2.10. Different forms of frequency distribution curves.

A *cumulative frequency distribution* is a tabulated form of cumulative frequencies according to the class intervals of the scores in the grouped data of a sample. For a cumulative frequency distribution (less-than type), the true upper limit (X_u) of each class interval is computed and recorded, because the *cf* of an interval includes in such a case the frequency of that interval too (Table 2.11).

Cumulative percentage

The cumulative percentage (*cP*) of a class interval is its cumulative frequency (*cf*), upto its X_u , expressed as a percentage of the total frequency (*n*) of the sample.

$$cP = \frac{cf}{n} \times 100.$$

A *cumulative percentage distribution* is a tabulated form of *cP* scores according to the class intervals of the scores in the grouped data of a sample (Table 2.11). For a *cP* distribution, the true upper limit (X_u) of each class interval is computed and recorded, because the *cP* of an interval takes into consideration the frequency of that interval also.

Cumulative frequency ogive

The *cf ogive* is the graphical representation of

the cumulative frequencies in the class intervals of the sample. To plot it for the less-than type, the scores (*X*) are scaled along the X axis, marking the true upper limits (X_u) of all the class intervals on that axis; in addition, the true lower limit (X_l) of the lowest class interval is also marked on the X axis as the X_u of the next lower interval with the *cf* of 0. The *cf* of each interval is plotted against its X_u and the plotted points are joined by straight lines to give the *cf ogive* of the less-than type. (Fig. 2.11).

Cumulative percentage ogive

The *cP ogive* is the graphical representation of the *cP* scores in the class intervals of the sample. To plot it, the scores (*X*) are scaled along the X axis, marking on the latter the true upper limits (X_u) of all the class intervals, as well as the X_l of the lowest interval as the X_u of the next lower interval with the *cP* of 0. The *cP* of each interval is plotted against its X_u and the plotted points are joined by straight lines to give the *cP ogive* (Fig. 2.12). *cP ogives* are used in the graphical methods for determining percentiles, percentile ranks, quartiles and median of a sample.

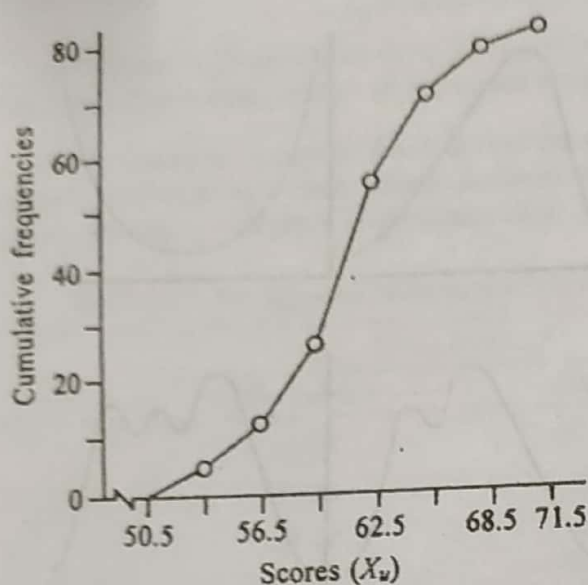


Fig. 2.11. Cumulative frequency ogive for body weights of a sample.

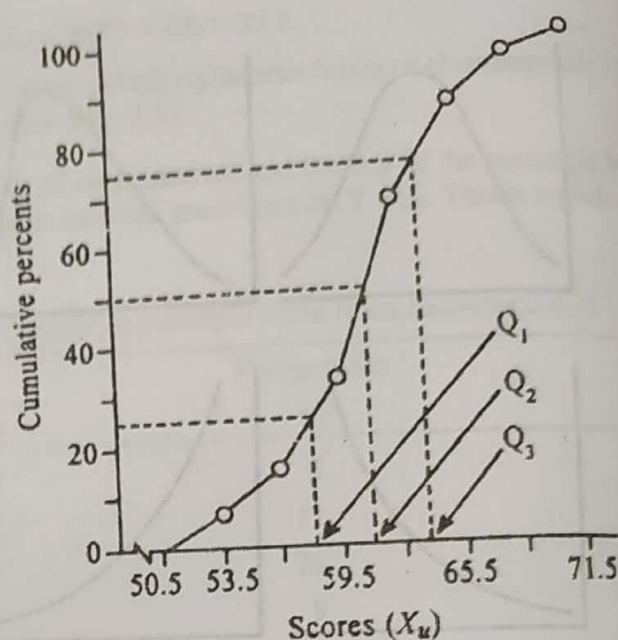


Fig. 2.12. cP ogive for body weights of a sample.

Example 2.7.1.

Draw the *cf* ogive and the *cP* ogive of the following distribution of body weights (kg) in a sample.

Class intervals :	51-53	54-56	57-59	60-62	63-65	66-68	69-71
Frequencies :	5	7	14	28	15	8	3

Solution :

1. The *cf* and *cP* distributions are first worked out as follows.

(a) The class intervals and their frequencies (*f*) are tabulated in Table 2.11. The empty class interval, immediately below the lowest interval containing observed frequencies, is also included in the table.

(b) The true upper limits (X_u) of all the intervals are computed and recorded. For example, for the interval 54-56,

$$X_u = \frac{1}{2} [(\text{upper score limit of interval}) + (\text{lower score limit of the next higher interval})]$$

$$= \frac{1}{2}(56 + 57) = 56.5.$$

(c) The cumulative frequency (*cf*) upto the X_u of each interval is then computed and entered in Table 2.11.

$$cf_k = f_1 + f_2 + f_3 + \dots + f_k$$

where cf_k is the *cf* upto the true upper limit of the *k*th interval from the lowest one, and f_1, f_2, \dots, f_k are the observed frequencies of the respective class intervals. For example, for the interval 54-56,

$$cf_3 = f_1 + f_2 + f_3 = 0 + 5 + 7 = 12.$$

(d) Each *cf* is converted to the corresponding *cP* and the computed *cP*s are entered in Table 2.11. For example, for the interval, 57-59,

$$cP = \frac{cf}{n} \times 100 = \frac{26}{80} \times 100 = 32.5.$$

Table 2.11. cf and cP distributions of body weights in a sample.

Class intervals	X_u	f	cf	cP
48-50	50.5	0	0	0.0
51-53	53.5	5	5	6.3
54-56	56.5	7	12	15.0
57-59	59.5	14	26	32.5
60-62	62.5	28	54	67.5
63-65	65.5	15	69	86.3
66-68	68.5	8	77	96.3
69-71	71.5	3	80 (n)	100.0

2. The cf ogive is then drawn as follows :

(a) The scores (X) are scaled along the X axis on a graph paper, marking the X_u of each class interval on that axis (Fig. 2.11).

(b) The cumulative frequencies (cf) are scaled along the Y axis.

(c) Each computed cf is plotted against the X_u of the corresponding interval, and all the plotted points are joined by straight lines to give the cf ogive.

3. The cP ogive is next drawn as follows :

(a) The X scores are scaled along the X axis of a graph paper, marking the X_u of each class interval on that axis (Fig. 2.12).

(b) The cumulative percentages (cP) are scaled along the Y axis.

(c) Each computed cP is plotted against the X_u of the corresponding interval and all the plotted points are joined by straight lines to give the cP ogive.

2.8 SCATTERGRAM

A combined frequency distribution of two variables is called a *bivariate frequency distribution*. One form of its graphical representation is the scattergram.

For a scattergram, the scores (X and Y) of two measurement variables are scaled along the abscissa and the ordinate, respectively. The Y score of each individual of the sample is then plotted against the X score of the same individual.

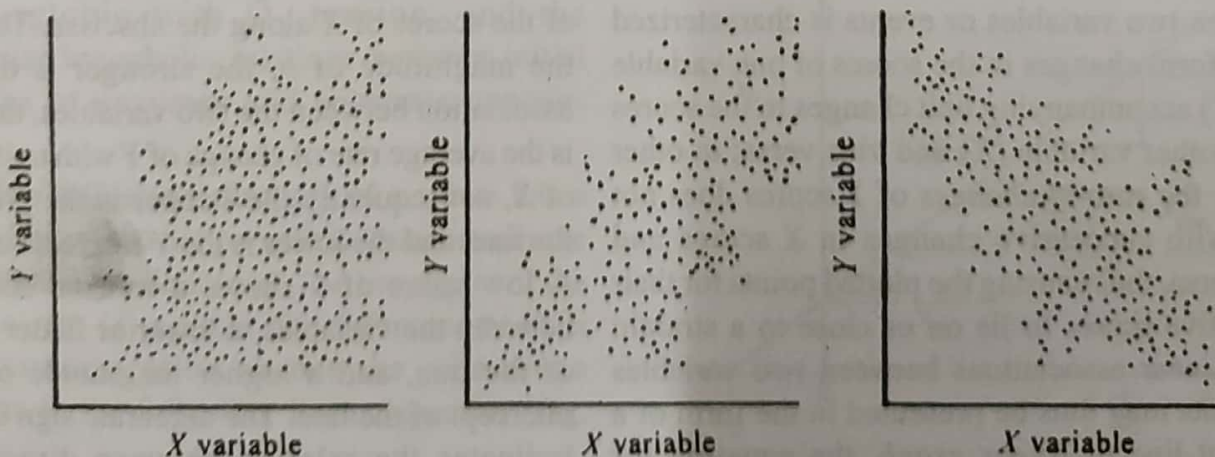


Fig. 2.13. Scattergrams of different types.

Each individual is thus plotted as a point located by his scores of the two variables. The distribution of all such points forms the scatter diagram (Fig. 2.13).

The scattergram may indicate the form of relationship between the two measurement variables. An irregular distribution of the plotted points may indicate the absence of any relationship between the variables. An elliptical or hyperbolic distribution of the scattergram indicates the possibility of a nonlinear association between the variables. If the plotted points are distributed around a straight line with a downward or upward slope, the scattergram indicates a negative or positive linear association between the variables.

2.9 LINEAR GRAPH

Where the magnitude (value or score) of one variable or event has some form of association with that of another, their magnitudes are dependent on those of each other, and each such variable or event may, therefore, be considered to be a *function* of the other. Frequently, two variables or events (X and Y) may have such association with each other that plotting of their scores against each other yields a distribution of the plotted points on or closely around a straight line. Such an association between the two paired variables or events is called a *linear association* and each of them may be considered to be a *linear function* of the other. Such linear association between two variables or events is characterized by uniform changes in the scores of one variable (say, Y) accompanying unit changes in the scores of the other variable (X) and vice versa; in other words, the *rate of changes* of Y scores does not vary with successive changes in X scores and vice versa, thus causing the plotted points for their respective scores to lie on or close to a straight line. Linear associations between two variables or events may thus be presented in the form of a straight line or *linear graph*, the equation for which is commonly expressed in the following

slope-intercept form :

$$Y = a + bX,$$

where Y and X are the variables scaled respectively along the ordinate (Y -axis) and the abscissa (X -axis), a is the intercept of the line with the Y -axis, and b is the slope of the line. Location of any point on the line is given or expressed by the *coordinates* or perpendicular distances of the point from respectively the X - and Y -axes. The slope-intercept equation changes into the following form where the straight line passes through the origin of the two axes and so, through the zero value of Y -intercept (i.e., $a = 0$) : $Y = bX$.

The term a is the *Y -intercept* of the line in its slope-intercept form and its magnitude varies opposite to that of the slope of the line. The magnitude as well as the algebraic sign of the term a indicates the general level of the line on the Y -axis. It indicates the expected, estimated or extrapolated magnitude of the variable Y for such a case as might have the zero magnitude of the variable X . A negative value of the term a signifies that the line has its Y -intercept located below the zero point of the ordinate scale.

The term b in the slope-intercept equation is the measure of *slope* of the line — the magnitude and the algebraic sign of the term b indicate respectively the steepness (*gradient*) and the *direction* of the slope. The term b gives a measure of the *average rate of change* of the scores of variable Y , scaled along the Y -axis, for unit change of the scores of X along the abscissa. The higher the magnitude of b , the stronger is the linear association between the two variables, the greater is the average rate of change of Y with unit change of X , consequently the steeper is the gradient of the line, and the lower is the Y -intercept or term a . A low value of b shows a weaker association between the variables, a lower or flatter gradient of the line, and a higher magnitude of the Y -intercept of the line. The algebraic sign of term b indicates the relation between directions of changes of the two variables. A *positive* b signifies

that high values of one variable are generally accompanied by high values of the other too, while low values of the former are associated with low values of the latter; in such a case, the line has an ascending gradient. On the contrary, a *negative b* shows that high and low values of one variable are generally associated with respectively low and high values of the other variable; thus, the line has a descending gradient and its slope-intercept equation is given by $Y = a - bX$.

Often the points, plotted with the observed paired scores of two variables in a sample, are so scattered that they cannot be joined directly to form a straight line. In such cases, the *best-fitting straight line* is so drawn as to keep the sum of squared vertical distances of the plotted points from that line at the lowest. This ensures the minimum total of the squared differences between the plotted points for the paired observations and the respective points for corresponding estimated values.

Examples of some linear plots are cited below. Mention should be made in this context that to gain greater precision in interpretations, extrapolations and estimations, equations for nonlinear associations are frequently transformed into linear equations for presentation as straight lines, such as the exponential relationships of radioactive decay with time and of vapour pressure with temperature, the sigmoid associations of initial velocities of allosteric enzyme actions with substrate concentrations and of oxygen saturation of hemoglobin with O_2 tension, and the rectangular hyperbolic relations between initial velocities of enzymes and substrate concentrations.

(a) The diffusional flux (J) of a solute has a positive linear association with its transmembrane concentration gradient (Δn). The linear plot of the diffusion rate (J) against the difference (Δn) in solute concentration across the membrane conforms to the following linear equation (Fig. 2.14):

$$J = P\Delta n, \text{ or, } J = (0+) P\Delta n,$$

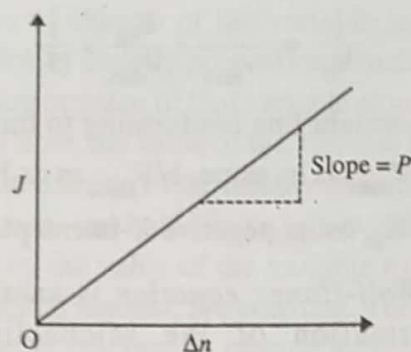


Fig. 2.14. Linear plot of diffusional flux (J) against transmembrane solute concentration (Δn).

where P is the permeability coefficient of the membrane for the solute in cm/sec for unit concentration difference and constitutes the slope of the line, while the Y -intercept of the latter coincides with the zero value of the diffusion rate on the ordinate scale.

(b) *Lineweaver-Burk double-reciprocal equation* is a linear transformation of the Michaelis-Menten rectangular hyperbolic equation for enzyme kinetics and is derived from the reciprocal of the latter equation. It expresses the reciprocal ($1/V_0$) of the initial velocity of enzyme action as a linear function of the reciprocal ($1/[S]$) of the molar concentration of the substrate, where V_{\max} is the maximum velocity of enzyme action and K_m is the Michaelis constant or the substrate concentration for attaining $\frac{1}{2}V_{\max}$ (Fig. 2.15).

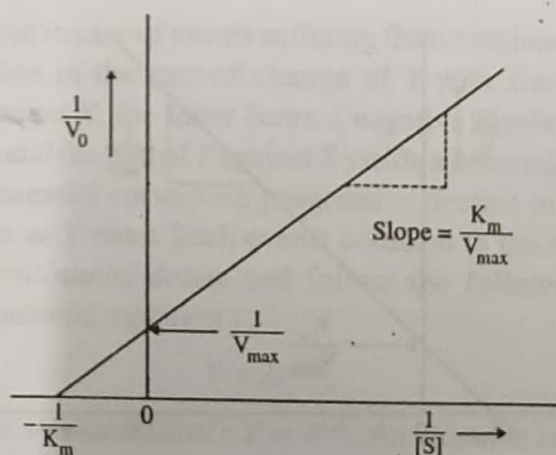


Fig. 2.15. Double-reciprocal plot.

$$\frac{1}{V_0} = \frac{1}{V_{\max}} + \frac{K_m}{V_{\max}} \times \frac{1}{[S]}.$$

The straight line conforming to this equation has K_m/V_{\max} as its slope, $1/V_{\max}$ as its Y -intercept, and $-1/K_m$ as its negative X -intercept.

(c) *Wolf-Hanes equation* is another linear transformation of the Michaelis-Menten hyperbolic equation, and is derived by multiplying both sides of the double-reciprocal equation by $[S]$. It expresses the $[S]/V_0$ ratio as a linear function of $[S]$ (Fig. 2.16).

$$\frac{[S]}{V_0} = \frac{K_m}{V_{\max}} + [S] \times \frac{1}{V_{\max}}.$$

The straight line conforming to this equation has $1/V_{\max}$ as its slope, K_m/V_{\max} as its Y -intercept, and $-K_m$ as its negative X -intercept.

(d) *Hill equation* for the sigmoid kinetics of allosteric enzymes is logarithmically transformed into the following equation expressing $\log [V_0/(V_{\max} - V_0)]$ as a linear function of $\log [S]$. Where K' is a constant,

$$\log \left[\frac{V_0}{V_{\max} - V_0} \right] = n \log [S] - \log K'.$$

The straight line resulting from this equation has as its slope the Hill coefficient (n) which indicates the type and magnitude of cooperativity

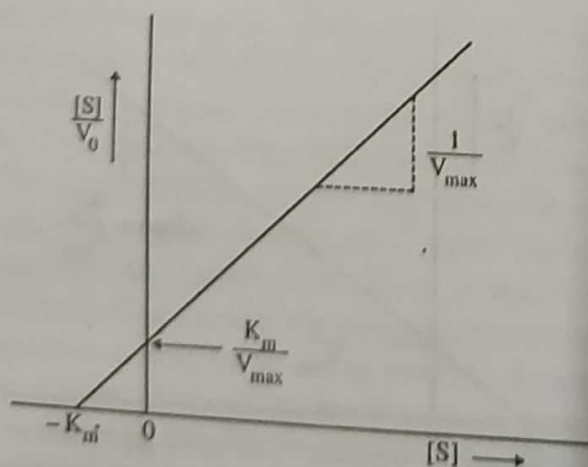


Fig. 2.16. Wolf-Hanes plot.

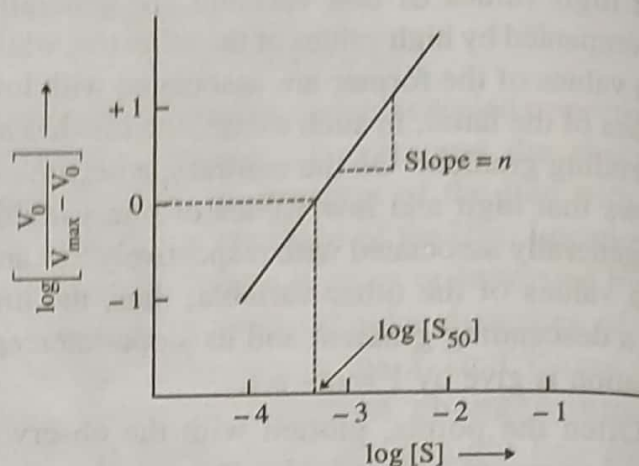


Fig. 2.17. Linear logarithmic plot of Hill equation.

between ligands binding to the enzyme. The Y -intercept of the line is $-\log K'$ (Fig. 2.17).

(e) The positive exponential relation between the vapour pressure (p) of a liquid and its absolute temperature (T) may be transformed logarithmically into the following linear equation expressing $\log p$ as a negative linear function of $1/T$.

$$\log p = C - \frac{L_v}{2.303R} \times \frac{1}{T},$$

where C is a constant depending on the liquid, R is the molar gas constant, and L_v is the molar heat

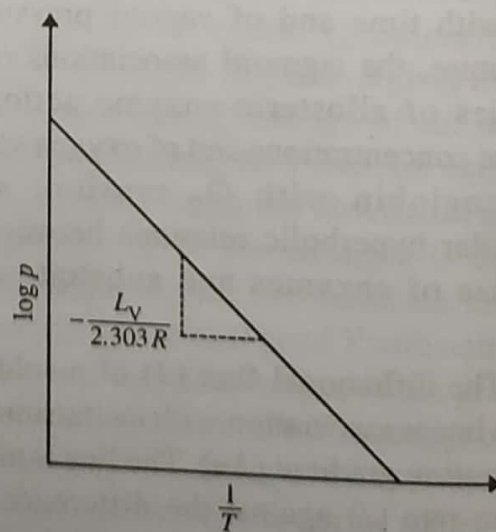


Fig. 2.18. Linear relationship between the logarithm of vapour pressure (p) and the reciprocal of absolute temperature (T).

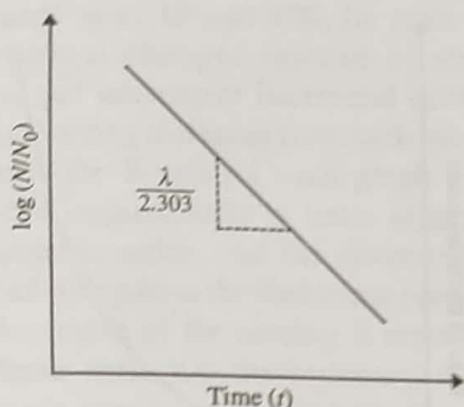


Fig. 2.19. Linear relationship between $\log (N/N_0)$ and time t .

of vaporization absorbed in vaporizing 1 mole of the liquid. The plotted line has a descending slope of $-L_v/2.303R$ and the Y -intercept equalling C (Fig. 2.18).

(f) The exponential equation for the radioactive decay of the nuclei of a specific radioisotope in a time interval (t) is transformed into the following equation for a straight line with a descending slope of $-\lambda/2.303$ where λ is the decay constant of that radioisotope while N_0 and N are the numbers of radioisotope nuclei existing respectively at the commencement and the termination of the interval t (Fig. 2.19).

$$\log \frac{N}{N_0} = -\frac{\lambda}{2.303}t = 0 - \frac{\lambda}{2.303}t.$$

(g) The linear equation of association between two variables can be used in working out the *linear regression* for predicting the likely score of either of them from the observed score of the other in an individual (see §8.12).

2.10 EXPONENTIAL CURVE

Where a process or event consists of changes of a single variable and its rate or velocity (V) depends on the amount or concentration $[A]$ of that variable alone, it is called a *first-order reaction* with k as its rate constant.

$$V = k[A].$$

The rate of change of the variable in a first-order reaction is directly proportional to the initial concentration or value of that variable alone. Thus, in each unit time, the value of the variable changes by a constant fraction of its initially existing value. In such events, the half-life ($t_{0.50}$), which is the time taken by the value of the variable to change by half its initial amount, is a constant. This makes the variable or reactant to change *exponentially* with time; in other words, an almost infinitely long time interval would be required in attaining the change of the total amount of the initially existing variable. An equation of such a change expresses the changing variable (Y) as a function of an *exponent* or power (X) of the base (e) of natural logarithm and assumes the following alternative forms, according as the exponent bears respectively positive and negative algebraic signs and depending on the constants Z and k .

Where the exponent (X) bears a positive algebraic sign and the plotting of Y and X scores against each other yields an *ascending exponential curve* with a progressively steeper slope as X increases,

$$Y = Ze^{kX}.$$

This applies to events complying with the *law of continuous growth*, having progressively increasing rates of changes of the variable Y with the rise in the exponent X . The standard form of this equation used generally is : $Y = e^X$.

But in case of events suffering from continuous decline in the rate of change of Y with rise of exponent X , the latter bears a negative algebraic sign and the plot of Y against X yields a descending exponential curve with progressive decline in its slope as X rises. Such events conform to the *law of continuous decay* and follow the following exponential equation :

$$Y = Ze^{-kX},$$

or its standard form : $Y = e^{-X}$. An example of a descending exponential curve is that for the continuous decline in the relative activity of a

radioisotope with time. This constitutes a first-order reaction with its rate directly proportional to only the initially existing number (N_0) of radioactive nuclei and changes exponentially with time complying with the law of decay. When N is the number of radioactive nuclei left after a time interval (t) and λ is the decay constant consisting of that fraction of radioactive nuclei which decays in each unit time interval,

$$Y = e^{-kX}, \text{ or } \frac{N}{N_0} = e^{-\lambda t}.$$

So, plotting of $\frac{N}{N_0}$ against t gives a negative exponential curve with its slope getting progressively less steep with time (Fig. 2.20).

On the contrary, the continuous rise in bacterial growth rate in a bacterial culture with time can be expressed as an ascending exponential curve (Fig. 2.21). This constitutes a first-order reaction with its rate directly proportional to the existing bacterial count, so as to rise exponentially with time in compliance with the law of continuous growth and following the standard equation : $Y = e^X$. The curve has an upward slope whose

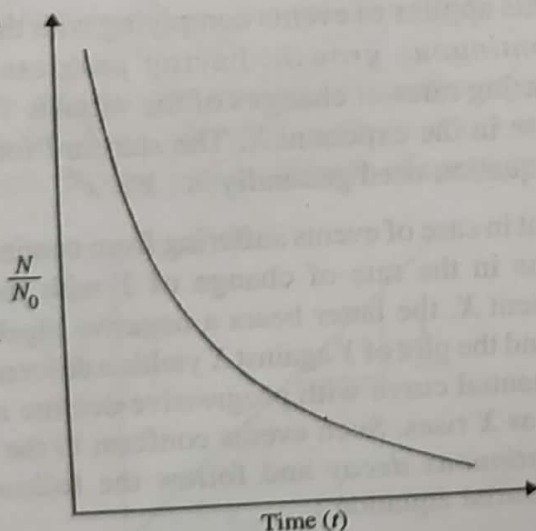


Fig. 2.20. Descending exponential curve ($Y = e^{-X}$) of the decline of relative activity of radioisotope with time.

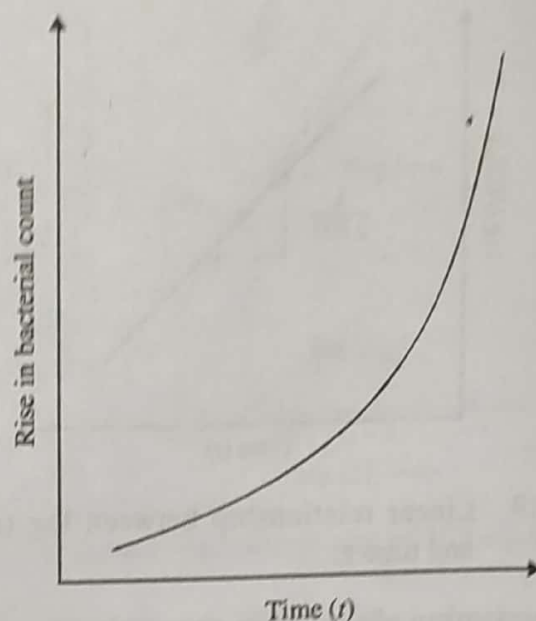


Fig. 2.21. Ascending exponential curve ($Y = e^X$) of rise in bacterial count in bacterial cultures with time.

gradient becomes progressively steeper with the length of time interval ($X = t$) and the Y -intercept is at the value 1 of the ordinate scale.

Exponential equations may be logarithmically transformed into linear equations and presented as straight line graphs on natural or ordinary graph papers because logarithms of numbers of any geometric series constitute an arithmetic series (see §2.9). But instead of such linear forms attained by using $\log Y$, the score Y itself of an exponential series may be used directly in drawing a semi-logarithmic graph.

2.11 SEMI-LOGARITHMIC GRAPH

A natural or ordinary graph paper has both its vertical (ordinate) and horizontal (abscissa) axes scaled arithmetically. So, successive rulings are equidistant from each other in both horizontal and vertical directions on it, and both axes may start from zero values of their respective scales. On a *semi-logarithmic graph paper*, although the abscissa (X -axis) is scaled arithmetically starting from the zero point and successive vertical rulings are equidistant from each other, the ordinate

(Y-axis) is scaled logarithmically from a non-zero origin such as 1, 10 and 100, its scale shows proportionate changes instead of absolute changes, and successive horizontal rulings are located at varying distances from each other. This is because the Y-axis on such graph paper is subdivided longitudinally in terms of geometric or logarithmic series, and the distance of each horizontal ruling from the abscissa is proportional to the logarithm of the number it represents on the ordinate scale. On the contrary, a *double-logarithmic graph paper* has both the abscissa (X-axis) and the ordinate (Y-axis) subdivided in terms of some geometric or logarithmic series so that neither horizontal nor vertical rulings are equidistant, the scale of neither axis starts from zero, and both scales show proportionate changes instead of absolute changes.

Where a variable Y has an exponential relationship with the power X of a base (say, e of natural logarithm) in terms of an equation such as $Y = e^X$, the Y and X scores may be scaled along the ordinate and the abscissa respectively, and each observed Y score — instead of $\log Y$ — may be directly plotted against the corresponding value

of X , thus yielding a linear *semi-logarithmic graph* with the plotted points. Such a graph is also known as the *ratio chart*, because the Y values form a geometric series having a constant ratio between successive scores or changing by a constant proportion. If the plotted points fall on the line or occur close around the latter, a constant rate prevails for the changes of Y variable; dispersions of the points away from the line indicate wider deviations from a constant rate.

Variations in the slope of the semi-logarithmic graph signify alterations in the rate of changes of Y variable. Rates of changes of two or more series of scores in geometric or exponential progressions may be compared using their respective semi-log graphs; semi-logarithmic lines would have parallel slopes if the corresponding series of scores have an identical rate of changes while lines with different slopes would indicate differences between different series in their rates.

Semi-log graphs can compare also two or more geometric or exponential series of scores in different units. They are also needed when the Y scores constitute a series with very wide range of values.

GLOSSARY

bar diagram : a set of parallel equal-width bars, separated from each other and having their areas proportional to the frequencies of cases or the amounts in different classes of a discontinuous or nominal variable in the sample.

bar diagram, multiple : several sets of parallel, separated and equal-width bars for as many samples, each set with bars having areas proportional to the frequencies of cases or the amounts in different classes of a discontinuous or nominal variable in one of the samples.

bar diagram, proportional : a set of parallel and separated bars of identical widths and lengths, each bar divided into several lengthwise segments with their lengths proportional to the frequencies of cases or the amounts in different classes of a variable in one of the several samples.

bar diagram, simple : a single set of parallel, separated and equal-width bars with their lengths proportional to the frequencies of cases or the amounts in different classes of a discontinuous or nominal variable in a single sample.

exponential equation : an equation expressing the values of a variable as a function of the power of a base such as the base of natural logarithm.

frequency : the number of repeated occurrences of a score or case in the sample.

- frequency, cumulative** : the total frequency of scores or cases in all the class intervals from one end of the frequency distribution of a sample upto the lower or upper true limit of a given interval.
- frequency distribution** : a distribution of frequencies of cases or scores of a sample in different class intervals of a variable.
- frequency polygon** : a polygon drawn as a graphical representation of the frequency distribution of a continuous variable by plotting the frequencies of scores in different classes of the variable against the midpoints of the respective classes.
- frequency table, simple** : an ungrouped frequency distribution of a discrete variable, in which the frequencies of different scores of a sample are entered against the respective individual scores, not in class intervals consisting of the respective ranges of scores.
- histogram** : the graphical representation of the frequency distribution of a continuous variable in a sample in the form of a set of equal-width bars without intervening gaps, each bar with its base extending between the true class limits of a class interval.
- linear equation** : an equation expressing the values of a variable as a linear function of another.
- ogive** : a graph drawn by plotting either the cumulative frequencies or the cumulative percentages of the classes of a variable in a sample against the true upper limits of the respective class intervals.
- percentage, cumulative** : cumulative frequency of any class interval expressed as a percentage of the sample size.
- pie diagram** : a circle with its different segments having their respective areas proportional to the relative frequencies of cases in different classes of a nominal variable.
- scattergram** : a graphical distribution of points drawn by plotting the scores of two particular variables of each individual or case of a sample against each other to study the form of association between those variables.
- semi-logarithmic graph paper** : graph paper with the abscissa scaled arithmetically making the vertical rulings equidistant, but with the ordinate scaled logarithmically resulting in unequal distances between successive horizontal rulings.

3. STATISTICS OF LOCATION

Statistics of location belong to the class of *descriptive statistics* (page 11). They serve to locate specific positions of the frequency distribution of a variable in a sample on the scale of scores of that variable. They include mean, median, mode, percentiles, deciles and quartiles.

3.1 CLASSIFICATION

Statistics of location are classified further into two main classes.

(a) Measures for central values :

They include mean, geometric mean, median and mode, and are also called *central tendencies*. They describe the locations of specific central positions of the frequency distribution of a variable in a sample on the scale of that variable.

(b) Quantiles or fractiles :

They are the scores below which lie specific fractions of the frequency distribution of a variable in a sample. They thus partition out specified fractions like specific numbers of quarters, one-tenths and one-hundredths of the distribution. They are also called *partition values* and include quartiles, deciles and percentiles. Median is both a measure of central value and a partition value.

3.2 MEAN

Mean is the arithmetic average of a set of scores. The mean of a sample (statistical mean) and that of a population (parametric mean) are represented by the symbols \bar{X} and μ , respectively. Where X (or X_i) represents each individual score of a sample, ΣX (or ΣX_i) is the sum of all its scores, and n is the sample size or the total frequency of cases in the sample,

$$\bar{X} = \frac{\Sigma X}{n}; \quad \text{or, } \bar{X} = \frac{\Sigma X_i}{n}.$$

This is also how the mean is worked out for a small sample, with its scores few in number and not arranged in a frequency distribution or a frequency table. Properties of mean are as follows:

(a) The sum of all the scores of a sample is given by the product of their mean and the sample size.

$$\Sigma X = n\bar{X}; \quad \text{or, } \Sigma X_i = n\bar{X}.$$

(b) Mean would be the score of each individual if the total score of the sample (ΣX) were equally distributed among all its individuals.

(c) The sum of positive deviations of some of the scores from the mean equals that of negative deviations of the remaining scores of the sample from it. So, the algebraic sum of the deviations of all the individual scores from the mean amounts to zero in any sample.

$$\Sigma(X - \bar{X}) = 0; \quad \text{or, } \Sigma(X_i - \bar{X}) = 0.$$

(d) The *sum of squares* about the mean, i.e., $\Sigma(X - \bar{X})^2$ or $\Sigma(X_i - \bar{X})^2$, is the sum of squared deviations of all the scores of a sample from its mean and is the *lowest* of all such sums of squares about the respective measures for central values.

(e) If the individual scores of a sample are all multiplied or divided by a constant number (say, k), the mean also gets respectively multiplied or divided by the same number.

$$\frac{\Sigma(kX)}{n} = k\bar{X}; \quad \frac{\Sigma X}{kn} = \frac{\bar{X}}{k}.$$

(f) If a constant number k is added to or subtracted from each score of a sample, its mean also gets respectively increased or decreased by the same number.

$$\frac{\Sigma(X + k)}{n} = \bar{X} + k; \quad \frac{\Sigma(X - k)}{n} = \bar{X} - k.$$

(g) In a bilaterally symmetrical and unimodal distribution, with a single peak and neither of its tails longer or more tapering than the other, the mean is the exactly central score and identical with the median and the mode. For such a distribution of scores in a sample, mean is the most reliable, stable and widely applicable central value.

(h) The presence of a score, with an extreme positive or negative deviation from the mean and not counterbalanced by the presence of another score with an equal but opposite deviation, makes the distribution asymmetric, displaces the mean towards that extreme score, and causes the mean to differ from both the median and the mode of the distribution. In such cases, $\bar{X} > Mdn > Mo$, if the unbalanced extreme score/scores is/are in the high-value (positive) tail of the distribution to make that tail longer; on the contrary, $\bar{X} < Mdn < Mo$, if such extreme scores occur in the low-value (negative) tail making the latter longer. This implies that the mean is unreliable as a central value in an asymmetric distribution which has one tail longer than the other due to a few scores with larger deviations in the longer tail.

Example 3.2.1.

Compute the mean of the following interorbital width scores (mm) of a sample of pigeons.
12.8, 11.7, 12.3, 10.8, 12.5, 11.4, 12.5, 10.9, 11.6, 11.7.

Solution :

$$\bar{X} = \frac{\sum X}{n} = \frac{12.8 + 11.7 + 12.3 + 10.8 + 12.5 + 11.4 + 12.5 + 10.9 + 11.6 + 11.7}{10} = 11.8 \text{ mm.}$$

Example 3.2.2.

Compute the mean of following body weight scores (kg) in a sample of humans.

66, 50, 56, 63, 68, 60, 60, 68, 72, 66, 68, 63, 60, 66, 75, 56, 72, 63.

Solution :

The sample being a small one ($n = 18$) with only a few scores occurring repeatedly in it, the data are arranged in a simple frequency table (Table 3.1).

$$\bar{X} = \frac{\sum fX}{n} = \frac{1152}{18} = 64.0 \text{ kg.}$$

(i) If the scores (Y) of a variable are the linear functions of the scores (X) of another variable, then the mean \bar{Y} of the former is also a linear function of the mean \bar{X} of the latter. Thus, if a is the vertical intercept and b is the slope of the straight line formed by plotting the Y scores against the X scores of the respective individuals in a sample,

$$Y = a + bX, \text{ and}$$

$$\bar{Y} = a + b\bar{X}.$$

Computation from frequency tables

In simple frequency tables with frequencies entered against single distinct scores, each forming a class by itself (Table 3.1), one or more individuals possess identical scores so that the scores are repeated in the data; but the data are not classified into groups. The mean is computed here from the frequencies (repetitions) of the individual scores. Where f_1, f_2, \dots, f_k are the frequencies (f_i) of the respective individual scores (X_i) like X_1, X_2, \dots, X_k , and n is the total frequency or sample size,

$$\bar{X} = \frac{\sum f_i X_i}{n} = \frac{f_1 X_1 + f_2 X_2 + \dots + f_k X_k}{n}.$$

Table 3.1. Frequency table of body weight scores for computation of mean.

Scores (X)	Frequencies (f)	fX
50	1	50
56	2	112
60	3	180
63	3	189
66	3	198
68	3	204
72	2	144
75	1	75
Total	18 (n)	1152 (ΣfX)

Computation from grouped data

This method is used for directly computing the mean of a continuous measurement variable whose scores have been arranged into a regular frequency distribution.

The entire range of the observed scores is divided into class intervals, preferably of *equal sizes*, and the frequency of scores falling in each interval is entered against the latter to form a frequency distribution (see pages 15-16). Because the score of each observation in a class interval is assumed to be identical with the midpoint (X_c) of that interval, the sum of the scores of each interval is obtained as fX_c by multiplying the frequency (f) of that interval with its midpoint (X_c). Thus,

the sum of the fX_c values of all the intervals gives the net sum of all the scores of the sample. Hence,

$$\bar{X} = \frac{\text{sum of all scores}}{\text{sample size}} = \frac{\Sigma fX_c}{n}$$

Minor discrepancies may arise in the mean computed in this way if the scores are grouped in a different set of class intervals. Moreover, the mean computed by using the midpoints of intervals may differ slightly from that computed directly from individual scores of ungrouped data.

The mean cannot be computed in this way if the data have been arranged in an *incomplete distribution with open class intervals* (page 16) because the midpoint is not available for such an interval.

Example 3.2.3.

Compute the mean body weight from the following frequency distribution of body weights (kg) in a sample of humans.

Class intervals :	51-53	54-56	57-59	60-62	63-65	66-68	69-71
Frequencies :	4	7	12	25	13	6	3

Solution :

(i) The data are arranged in Table 3.2, and the midpoint (X_c) of each class interval is computed and entered in the table. For example, for the interval 54-56.

$$X_c = \frac{1}{2}[(\text{upper score limit}) + (\text{lower score limit})]$$

$$\text{or, } X_c = \frac{1}{2}(56 + 54) = 55.$$

Table 3.2. Table for computing the mean body weight from grouped data.

Class intervals	X_c	f	fX_c
51-53	52	4	208
54-56	55	7	385
57-59	58	12	696
60-62	61	25	1525
63-65	64	13	832
66-68	67	6	402
69-71	70	3	210
Total		70 (n)	4258

(ii) Each X_c is multiplied by the frequency (f) of cases in that interval to work out the fX_c of the latter. For example, for the interval 57-59, $fX_c = 12 \times 58 = 696$.

(iii) The sum of the fX_c values of all the intervals and the total frequency (n) of the sample are used to compute the mean \bar{X} .

$$\bar{X} = \frac{\sum fX_c}{n} = \frac{4258}{70} = 60.8 \text{ kg.}$$

Example 3.2.4.

Compute the mean for the wing lengths (mm) of houseflies given below :

3.9, 4.3, 4.8, 4.7, 4.6, 4.4, 3.7, 4.2, 4.1, 4.8, 5.3, 4.9, 4.6, 3.8, 4.0, 5.3, 5.7, 5.5, 3.9, 4.5,
3.8, 4.5, 5.0, 4.9, 4.8, 3.5, 4.3, 5.1, 3.9, 4.7, 5.6, 4.6, 4.4, 3.4, 5.1, 4.6, 3.9, 3.8, 4.8, 4.9.

Solution :

(i) The data are first arranged into a frequency distribution and entered in Table 3.3.

Highest score = 5.7.

Lowest score = 3.4.

Range = 3.4 to 5.7.

Sample size (n) = 40.

Number of intervals chosen = 5.

Size of class intervals (i) $[(5.7 - 3.4) \div 0.1] \div 5 \approx 0.5$.

(ii) The midpoint X_c is computed for each class interval. For example, for the interval 3.9 - 4.3,

$$X_c = \frac{1}{2}(4.3 + 3.9) = 4.1.$$

(iii) Each X_c is multiplied by the frequency (f) of that interval to give fX_c . For the interval 3.9 - 4.3, for example,

$$fX_c = 9 \times 4.1 = 36.9.$$

(iv) The sum of the fX_c values of all intervals and the total frequency (n) of the sample are used in computing the mean (\bar{X}).

$$\bar{X} = \frac{\sum fX_c}{n} = \frac{180.5}{40} = 4.51 \text{ mm.}$$

Table 3.3. Table for computing the mean winglength from grouped data.

Class intervals	X_c	f	fX_c
3.4-3.8	3.6	6	21.6
3.9-4.3	4.1	9	36.9
4.4-4.8	4.6	14	64.4
4.9-5.3	5.1	8	40.8
5.4-5.8	5.6	3	16.8
Total		40	180.5

Computation of weighted mean

using the group sizes (n_1, n_2 , etc.) as the weights for the respective group means.

The means (\bar{X}_1, \bar{X}_2 , etc.) of a given variable in k number of groups or samples may be used to compute the *weighted mean* (\bar{X}) of the full set,

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + \dots + n_k\bar{X}_k}{n_1 + n_2 + \dots + n_k}$$

Example 3.2.5.

The mean systolic blood pressure was found to be 129.4 and 133.6 mm Hg for two groups of 12 and 15 humans, respectively. Find the mean systolic blood pressure of all the 27 men.

Solution :

$$\bar{X}_1 = 129.4; \quad n_1 = 12. \quad \bar{X}_2 = 133.6; \quad n_2 = 15.$$

$$\therefore \bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2} = \frac{12 \times 129.4 + 15 \times 133.6}{12 + 15} = 131.7 \text{ mm Hg.}$$

Example 3.2.6.

20% of a group of 80 men and 15% of a group of 120 women were found to be diabetic. Find the mean percentage of diabetics for both the groups combined.

Solution :

$$\text{For men : } n_1 = 80; \text{ percentage } (P_1) = 20. \quad \text{For women : } n_2 = 120; \quad P_2 = 15.$$

$$\therefore \bar{X}_P = \frac{n_1P_1 + n_2P_2}{n_1 + n_2} = \frac{80 \times 20 + 120 \times 15}{80 + 120} = 17\%.$$

3.3 GEOMETRIC MEAN

Geometric mean (GM) is the n th root of the product of all scores (X) of a sample when n is the total number of scores and none of the scores is negative or 0. It is thus the antilog of the mean of the logarithms of scores, provided all of the scores are higher than 0 in value. Where Π

indicates the product of the terms following it,

$$GM = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \cdot \dots \cdot X_n} = (\Pi X)^{\frac{1}{n}}$$

$$= \text{Antilog} \left[\frac{\sum \log X}{n} \right].$$

GM is computed for measurements in a

logarithmic scale, frequency distributions with scores more concentrated in the low-value tail, and distributions of bacterial counts, reflex reaction times, and sensory responses like differential perception of sound frequencies of matched intensities and optical potentials at different illuminations. Some of its properties are as follows.

(a) The product of all the scores (X) of a sample of size n equals the n th power of GM .

$$\prod X = (GM)^n.$$

(b) The arithmetic average of the logarithms of all the scores of a sample (size = n) equals the

logarithm of their GM .

$$\log GM = \frac{\sum \log X}{n}.$$

(c) Where all scores have positive values, \bar{X} ordinarily either exceeds or equals GM .

(d) Where there are k number of samples with respective sizes of n_1, n_2, \dots, n_k , the weighted GM of all the samples is computed from their respective GM s and sample sizes.

$$\log GM = \frac{1}{n_1 + n_2 + \dots + n_k} (n_1 \log GM_1 + n_2 \log GM_2 + \dots + n_k \log GM_k).$$

Example 3.3.1.

Compute the geometric mean of the following pH values of a number of bacterial cultures :

8.0, 8.3, 7.1, 8.2, 7.6, 7.7, 7.9, 8.0, 7.4, 8.2, 8.4, 8.1, 7.8, 7.9, 7.6.

Solution :

Table 3.4. Treatment of pH data for GM .

Serial No.	pH scores (X)	$\log X$	Serial No.	pH scores (X)	$\log X$
1	8.0	0.9030899	Total brought forward		
2	8.3	0.919078	9	7.4	0.8692317
3	7.1	0.8512583	10	8.2	0.9138138
4	8.2	0.9138138	11	8.4	0.9242792
5	7.6	0.8808135	12	8.1	0.908485
6	7.7	0.8864907	13	7.8	0.8920946
7	7.9	0.897627	14	7.9	0.897627
8	8.0	0.9030899	15	7.6	0.8808135
Total carried forward		7.1552611	Total ($\sum \log X$)		
			13.4416059		

$$GM = \text{Antilog} \left[\frac{\sum \log X}{n} \right] = \text{Antilog} \left[\frac{13.4416059}{15} \right] = 7.87.$$

3.4 QUANTILES OR FRACTILES

These statistics of location include percentiles, quartiles and deciles. These partition values are used in working out quartile deviation, skewness and kurtosis (vide chapters 4 and 6).

Percentiles (P_p) are those scores in a frequency distribution, below which lie specific percentages of the total number of scores of the sample. Thus, the 50th percentile (P_{50}) is that score below which lies the lowest 50% of the scores ; P_{50} is identical

with Mdn (§ 3.6). Similarly, P_{25} and P_{75} are the scores below which lie respectively 25% and 75% of the total scores.

Quartiles are those scores in a frequency distribution, below which lie specific numbers of quarters of the total frequency. Thus, Q_1 , Q_2 , Q_3 and Q_4 are the scores below which lie respectively one-fourth, half, three-fourths and all of the total number of scores. $Q_2 = P_{50} = Mdn$; $Q_1 = P_{25}$; $Q_3 = P_{75}$; $Q_4 = P_{100}$.

Deciles are those scores in a frequency distribution, below which lie specific numbers of tenth parts of the total distribution. Thus, D_1 , D_3 and D_{10} , the first, third and tenth deciles, are the scores below which lie the lowest $\frac{1}{10}$ th, $\frac{3}{10}$ ths and

all of the scores respectively. D_5 is identical with Mdn , P_{50} and Q_2 , D_1 with P_{10} , and D_{10} with P_{100} .

Computation of quantiles

Where P_p is the required percentile, cf_l is the cumulative frequency of all the intervals below the true lower limit X_l of the class interval containing P_p , f_p is the observed frequency in that class interval, i is the interval size, n is the sample size, p is the proportion of total cases below P_p , and pn is the number of cases to be counted off in reaching P_p from the lowest score,

$$P_p = X_l + i \times \frac{pn - cf_l}{f_p}$$

Deciles, quartiles and median are likewise computed as the corresponding percentiles.

Example 3.4.1.

Compute the 25th and 75th percentiles of the frequency distribution of body weights (kg) of Table 2.11.

Solution :

The cumulative frequency distribution of Table 2.11 is reproduced in Table 3.5.

(a) The size i of the intervals is obtained by subtracting X_u of any interval from that of the next higher one. Using the X_u scores of 65.5 and 62.5.

$$i = 65.5 - 62.5 = 3.$$

(b) For computing P_{25} :

$$p = 0.25 ; \quad n = 80 ; \quad \therefore pn = 0.25 \times 80 = 20.$$

So, 20 scores are counted off starting from the lowest class interval, thus reaching into the interval 57-59 in which P_{25} lies. The true lower limit (X_l) and the frequency (f_p) of that interval, and the cumulative frequency cf_l upto its lower limit amount to 56.5, 14 and 12, respectively.

$$X_l = 56.5 ; \quad cf_l = 12 ; \quad f_p = 14 ; \quad pn = 20 ; \quad i = 3 ;$$

$$P_{25} \text{ or } Q_1 = X_l + i \times \frac{pn - cf_l}{f_p} = 56.5 + 3 \times \frac{20 - 12}{14} = 58.2 \text{ kg.}$$

Table 3.5. Cumulative frequencies of body weights (kg).

Class intervals	True limits		f	cf
	lower (X_l)	upper (X_u)		
51-53	50.5	53.5	5	5
54-56	53.5	56.5	7	12
57-59	56.5	59.5	14	26
60-62	59.5	62.5	28	54
63-65	62.5	65.5	15	69
66-68	65.5	68.5	8	77
69-71	68.5	71.5	3	80 (n)

(c) For computing P_{75} :

$$p = 0.75 ; \quad pn = 0.75 \times 80 = 60.$$

In counting off 60 scores starting from the lowest interval, the interval 63-65 is reached in which lies P_{75} .

$$X_l = 62.5 ; \quad cf_l = 54 ; \quad f_p = 15 ; \quad pn = 60 ; \quad i = 3 ;$$

$$P_{75} \text{ or } Q_3 = X_l + i \times \frac{pn - cf_l}{f_p} = 62.5 + 3 \times \frac{60 - 54}{15} = 63.7 \text{ kg.}$$

Graphical determination

Percentiles, deciles and quartiles may be obtained graphically from cP ogives (Fig. 2.12). A line is drawn parallel to the X axis from that point on the Y axis which corresponds to the cP for the required fractile ; e.g., the cP amounts to 10, 20, 25, 50 and 75 for respectively D_1 , D_2 , P_{25} , P_{50} and P_{75} . From the point of intersection of this line with the ogive, an ordinate is dropped to the X axis. The point of intersection of this ordinate with the X axis gives the required fractile.

cf_l upto X_l ,

$$PR = 100 \times \frac{cf_l + \frac{(X - X_l)f_p}{i}}{n}.$$

PR from ranked scores

(a) When the given score has been assigned numerical rank R in a descending order of magnitude in a sample of size n ,

$$PR = 100 - \frac{100R - 50}{n}.$$

(b) Where the ranking is in ascending order,

$$PR = \frac{100R - 50}{n}.$$

3.5 PERCENTILE RANKS

A percentile rank (PR) is the rank or graded position of a given score on a scale of 100 among all the scores of a sample. It is estimated from the percentage of scores lying below it.

PR from cumulative frequencies

Where the given score X belongs to a class interval having the size i , the frequency f_p , the true lower limit X_l and the cumulative frequency

PR from cP ogive

An ordinate is raised on the X axis of a cP ogive at the given score X . A horizontal line is drawn from the point of intersection of the ordinate with the ogive. The point of intersection of this line with the Y axis gives the PR of the given score.

Example 3.5.1.

Find the PR of the score 64 of the data presented in Table 3.5 of Example 3.4.1.

Solution :

The score 64 belongs to the interval 63-65 with the true lower limit (X_l) of 62.5, an interval size (i) of 3 and a frequency (f_p) of 15 ; the cumulative frequency (cf_l) upto the X_l of 62.5 amounts to 54 (Table 3.5).

$$X = 64 ; \quad X_l = 62.5 ; \quad i = 3 ; \quad f_p = 15 ; \quad cf_l = 54 ; \quad n = 80 ;$$

$$\therefore PR = 100 \times \frac{cf_l + \frac{(X - X_l)f_p}{i}}{n} = 100 \times \frac{54 + \frac{(64 - 62.5)15}{3}}{80} = 76.9.$$

Example 3.5.2.

Find the percentile ranks of students occupying the 3rd and 20th ranks in the descending order of merit in a Biology examination involving 80 students.

Solution :

(i) For the student with rank (R) of 3,

$$PR = 100 - \frac{100R - 50}{n} = 100 - \frac{100 \times 3 - 50}{80} = 96.9.$$

(ii) Similarly, for the student with rank (R) of 20,

$$PR = 100 - \frac{100 \times 20 - 50}{80} = 75.6.$$

3.6 MEDIAN

Median (Mdn) is that score in a frequency distribution, above and below which lie equal numbers, i.e., 50%, of the scores or cases of the sample. It is a partition value or fractile (§ 3.4) and is identical with D_5 , P_{50} and Q_2 . Some of its properties are given below.

(a) The ordinate on the X axis at the Mdn bisects the area of a frequency distribution into two equal halves.

(b) In symmetric unimodal distributions, (i) Mdn coincides with the mean and the mode, and (ii) the algebraic sum of deviations of the observed scores from the median amounts to 0.

$$Mdn = \bar{X} = M_0; \quad \Sigma(X - Mdn) = 0.$$

(c) In asymmetric distributions, (i) the median differs from the mean and the mode, the mean being located further than the median towards the longer-tail of the distribution, and (ii) the algebraic sum of deviations of the scores from the median differs from 0 in value, indicating by its positive or negative sign respectively a longer positive or negative tail of the distribution. Thus for an asymmetric distribution with a longer positive tail, $\bar{X} > Mdn > M_0$, and $\Sigma(X - Mdn) > 0$; but for a distribution with a longer negative tail, $M_0 > Mdn > \bar{X}$, and $\Sigma(X - Mdn) < 0$.

(d) As the median is less deflected than the

mean by extreme deviations of a few scores, it is a more reliable and representative measure of central value than the mean for an asymmetric distribution.

(e) Median can be computed even for frequency distributions with *open class intervals* or *unequal class sizes*, and also for ranked data like those of psychological and achievement tests.

Median is used in working out mean deviation, coefficient of mean deviation, coefficient of skewness and the median test.

Graphical determination

After drawing a cP ogive (Example 2.7.1, Fig. 2.12), a line is drawn parallel to its X axis from that point on the Y axis which corresponds to the cP of 50. From where this line meets the ogive, an ordinate is dropped to the X axis. The point of intersection of this ordinate with the X axis gives the Mdn .

Computation from ungrouped data

In an ungrouped set of data, median is the $(n + 1)/2$ th score, counted from either the lowest or the highest score of the sample.

(a) If there is an odd number of scores in the sample, i.e., n is an odd number, Mdn coincides with that observed score which belongs to the $(n + 1)/2$ th individual.

(b) If there is an even number of scores in the sample, the $(n+1)/2$ th score falls midway between two observed scores and is given by the average of those two scores (Example 3.6.1).

(c) If the *Mdn* or $(n+1)/2$ th score falls within a set of identical scores in the data, all the observed identical scores of that set are assumed to occupy one unit interval extending from 0.5 below the

score to 0.5 above the latter. Each score of the set is assumed to cover that fraction of this unit interval as is given by the reciprocal of the number of scores in that set. *Mdn* is computed by adding to the lower limit of this unit interval as many of these fractions of that interval as the number of identical scores of the set covered in counting off 0.50n scores, starting from the lowest score of the data (Example 3.6.2).

Example 3.6.1.

Find the median for the following reflex knee jerk strengths (in degrees of arc) of a sample of athletes :
19, 21, 22, 26, 28, 30, 31, 35, 35, 37.

Solution :

$$n = 10. \quad Mdn = \frac{n+1}{2} \text{th score} = \frac{10+1}{2} \text{ or } 5.5\text{th score.}$$

Thus, the *Mdn* lies between the 5th and 6th scores counted from either the lowest or the highest score. Counting from the lowest score of 19, the 5th and 6th scores amount to 28 and 30 respectively.

$$\therefore Mdn = 5.5\text{th score} = \frac{5\text{th score} + 6\text{th score}}{2} = \frac{28 + 30}{2} = 29^\circ.$$

Example 3.6.2.

Find the median for the following wing length (mm) data of a sample of cockroaches :
20, 22, 23, 24, 26, 26, 26, 28, 29, 29, 31.

Solution :

$$n = 11. \quad \therefore Mdn = \frac{n+1}{2} \text{th score} = \frac{11+1}{2} \text{ or } 6\text{th score.}$$

Thus, five scores lie below the *Mdn* and the remaining five above it. But in counting off five scores from the lowest one, the first of three identical scores, viz., the first 26, gets included in those five scores. However, the three identical scores forming the set of 26 may be assumed to occupy one unit interval extending from 25.5-26.5, each score occupying $\frac{1}{3}$ or 0.33 of this interval. On counting off only one of these three identical scores for arriving at the *Mdn*, the upper limit of that counted score is reached at $25.5 + 0.33$ or 25.83. This, therefore, is the *Mdn*. $\therefore Mdn = 25.8 \text{ mm.}$

Computation from grouped data

For a continuous frequency distribution, grouped into class intervals, *Mdn* is computed as P_{50} or Q_2 from cumulative frequencies.

$$Mdn = X_l + i \times \frac{0.50n - cf_l}{f_p}$$

where cf_l is the cumulative frequency of all class intervals below the true lower limit X_l of the

interval containing the Mdn , f_p is the observed frequency in that interval, i is the size of class intervals, and $0.50n$ gives the proportion of the total frequency n to be counted off from one end of the distribution to reach the Mdn .

Such computations may sometimes pose problems. (a) If the Mdn falls between

two intervals, the true limit between those two intervals, i.e., the true lower limit of the higher of the two intervals, is taken as the Mdn (Example 3.6.4). (b) If the Mdn falls in a vacant interval containing no case, the midpoint X_c of this interval is then taken as the Mdn (Example 3.6.5).

Example 3.6.3.

Compute the median of the following frequency distribution of body weights (kg).

Class intervals :	51-53	54-56	57-59	60-62	63-65	66-68	69-71
Frequencies :	5	7	14	28	15	8	3

Solution :

(a) The data are arranged in Table 3.6. The true lower limit (X_l) of each interval and the cumulative frequency (cf) upto its X_u are computed. For example, for the 4th interval 60-62 from the lowest one,

$$X_l = \frac{1}{2}[(\text{lower score limit of the interval}) + (\text{upper score limit of the next lower interval})] \\ = \frac{1}{2}(60 + 59) = 59.5.$$

$$cf = f_1 + f_2 + f_3 + f_4 = 5 + 7 + 14 + 28 = 54.$$

Table 3.6. Cumulative frequencies of body weight data. (cf upto the respective X_u scores)

Class intervals	X_l	f	cf
51-53	50.5	5	5
54-56	53.5	7	12
57-59	56.5	14	26
60-62	59.5	28 (f_p)	54 (median class)
63-65	62.5	15	69
66-68	65.5	8	77
69-71	68.5	3	80 (n)

(b) The size i of the class intervals is obtained by subtracting the X_l of any interval from the X_l of the next higher one. Thus, $i = 59.5 - 56.5 = 3$.

(c) The number of scores, to be counted off from one end of the distribution to reach the Mdn , is given by $0.50n$. Thus, $0.50n = 0.50 \times 80 = 40$.

(d) The counting off of 40 scores with effect from the lowest score leads into the interval 60-62 (median class) in which the Mdn lies. The true lower limit (X_l) of this interval is 59.5, the interval has a frequency (f_p) of 28 cases, and the cumulative frequency (cf_l) upto the (X_l) of this interval amounts to 26.

$$X_l = 59.5 ; \quad cf_l = 26 ; \quad f_p = 28.$$

(e) The Mdn is then computed as follows :

$$Mdn = X_l + i \times \frac{0.50n - cf_l}{f_p} = 59.5 + 3 \times \frac{40 - 26}{28} = 61.0 \text{ kg.}$$

Example 3.6.4.

Calculate the median for the following frequency distribution of achievement test scores in a group of students.

Class intervals :	67-76	77-86	87-96	97-106	107-116	117-126
Frequencies :	8	13	18	19	15	5

Solution :

The data are arranged in Table 3.7.

(a) The true lower limit (X_l) of each interval and the cumulative frequency (cf) upto its X_u are computed. For example, for the 4th interval 97-106 from the lowest,

$$X_l = \frac{1}{2}[(\text{lower score limit of the interval}) + (\text{upper score limit of the next lower interval})]$$

$$= \frac{1}{2}(97 + 96) = 96.5$$

$$cf = f_1 + f_2 + f_3 + f_4 = 8 + 13 + 18 + 19 = 58.$$

(b) The size i of the class intervals is obtained by subtracting the X_l of any interval from the X_l of the next higher one. Thus, $i = 86.5 - 76.5 = 10$.

Table 3.7. Cumulative frequencies of achievement test scores. (cf upto the respective X_u scores)

Class intervals	X_l	f	cf
117-126	116.5	5	78 (n)
107-116	106.5	15	73
97-106	96.5	19	58
87-96	86.5	18	39
77-86	76.5	13	21
67-76	66.5	8	8

(c) The number of scores, to be counted off from one end of the distribution to reach the Mdn , is given by $0.50n$. Thus, $0.50n = 0.50 \times 78 = 39$.

(d) The counting off of 39 scores, starting from the lowest interval, leads exactly upto the X_l of the interval 97-106. The Mdn , therefore, falls between the intervals 87-96 and 97-106. So, the X_l (96.5) of the higher of these two intervals, viz., 97-106, is taken as the median. Thus, $Mdn = 96.5$.

[The same result is also obtained on applying the formula used in the last example.

$$Mdn = X_l + i \times \frac{0.50n - cf_l}{f_p} = 96.5 + 10 \times \frac{39 - 39}{19} = 96.5.]$$

Example 3.6.5.

Calculate the median of the following frequency distribution of serum iron ($\mu\text{g dL}^{-1}$) in 32 humans.

Class intervals :	97-106	107-116	117-126	127-136	137-146	147-156
Frequencies :	3	5	8	0	11	5

Solution :

The data are arranged in Table 3.8.

(a) The true lower limit (X_l) of each interval and the cumulative frequency (cf) upto its X_u are computed. For example, for the 3rd interval 117-126 from the lowest one,

$$X_l = \frac{1}{2}[(\text{lower score limit of the interval}) + (\text{upper score limit of the next lower interval})]$$

$$= \frac{1}{2}(117 + 116) = 116.5.$$

$$cf = f_1 + f_2 + f_3 = 3 + 5 + 8 = 16.$$

(b) The size i of the class intervals is obtained by subtracting the X_l of any interval from the X_l of the next higher interval ; thus, $i = 116.5 - 106.5 = 10$.

Table 3.8. Cumulative frequencies of serum iron data. (cf upto the respective X_u scores)

Class intervals	X_l	f	cf
97-106	96.5	3	3
107-116	106.5	5	8
117-126	116.5	8	16
127-136	126.5	0	16 (median class)
137-146	136.5	11	27
147-156	146.5	5	32 (n)

(c) The number of scores, to be counted off from one end of the distribution to reach the Mdn , is given by $0.50n$. Thus, $Q_{50}n = 0.50 \times 32 = 16$.

(d) The counting off of 16 scores starting from the lowest interval brings us exactly upto the X_l of the interval 127-136, in which the Mdn should fall ; but this is a vacant interval with no case. So, the midpoint (X_c) of that interval is taken as the Mdn . For this interval,

$$X_c = \frac{1}{2}(X_u + X_l) = \frac{1}{2}(136.5 + 126.5) = 131.5. \quad \therefore Mdn = 131.5 \mu g.$$

3.7 MODE

The mode (M_o) is that score of the variable which belongs to the largest number of individuals in a sample. It is, therefore, the most frequent score in the sample and coincides with that point on the X axis of a frequency distribution which corresponds to the peak of the latter. Some of its properties are as follows.

(a) A distribution may be *unimodal*, *bimodal* or *multimodal*, according to its one, two or more peaks and as many M_o values.

(b) There is no mode if all scores of the sample are either identical or have the same frequency.

(c) In a perfectly symmetric unimodal distribution, M_o , \bar{X} and Mdn are identical.

(d) Mode, unlike median and mean, does not

change even if some extreme scores occur in only one tail of the frequency distribution.

(e) In an asymmetric unimodal distribution, Mdn lies between M_o and \bar{X} while M_o lies on that side of the Mdn which leads to the shorter tail of the distribution. Thus, for a distribution with a longer positive (high-value) tail, $\bar{X} > Mdn > M_o$; but if the negative (low-value) tail is longer than the positive one, $M_o > Mdn > \bar{X}$.

(f) The amount and algebraic sign of the deviation of the mean from the mode indicate respectively the degree and direction of asymmetry of the distribution.

Computations of mode

(a) In a simple series of scores or in the ungrouped data of a simple frequency distribution,

M_o is the most frequent score.

(b) In the grouped data of a quantitative frequency distribution with class intervals of equal size, the mode occurs in the interval having the highest frequency of scores. Where X_l is the true lower limit of the modal class having the highest frequency of scores and carrying the mode, f_m is the frequency of that modal class, f_{m-1} and f_{m+1} are the frequencies of the class intervals that respectively precede and follow the modal class and i is the size of each class interval,

$$d_1 = f_m - f_{m-1}; \quad d_2 = f_m - f_{m+1};$$

$$\therefore d_1 + d_2 = 2f_m - f_{m-1} - f_{m+1};$$

$$M_o = X_l + i \times \frac{d_1}{d_1 + d_2}.$$

M_o can be so computed even in incomplete frequency distributions with open class intervals.

(c) In a grouped distribution with class intervals of unequal sizes (lengths), mode is worked out approximately from the mean and the median of the sample : $M_o = 3Mdn - 2\bar{X}$. For example, the mode of a frequency distribution having a mean of 73.12 and a median of 73.0, is given by :

$$\begin{aligned} M_o &= 3Mdn - 2\bar{X} \\ &= 3 \times 73 - 2 \times 73.12 = 72.76. \end{aligned}$$

Example 3.7.1.

Compute the mode of the following frequency distribution of body weights (kg).

Class intervals :	51-53	54-56	57-59	60-62	63-65	66-68	69-71
Frequencies :	5	7	14	28	15	8	3

Solution :

(a) The data are arranged in Table 3.9. The true lower limit (X_l) of each interval is computed and entered there. For example, for the interval 60-62,

$$\begin{aligned} X_l &= \frac{1}{2}[(\text{lower score limit of the interval}) + (\text{upper score limit of the next lower interval})] \\ &= \frac{1}{2}(60 + 59) = 59.5. \end{aligned}$$

Table 3.9. Frequency distribution of body weight scores.

Class intervals	X_l	f
51-53	50.5	5
54-56	53.5	7
57-59	56.5	14
60-62	59.5	28 (modal class)
63-65	62.5	15
66-68	65.5	8
69-71	68.5	3

(b) The size i of the class intervals is obtained by subtracting the X_l of any interval from the X_l of the next higher interval ; thus, $i = 59.5 - 56.5 = 3$.

(c) The class interval 60-62, having the highest frequency of 28, is identified as the modal class. The frequencies of the modal class, the class preceding it and the class following it, are identified as f_m , f_{m-1} and f_{m+1} respectively and noted. Thus, $f_m = 28$; $f_{m-1} = 14$; $f_{m+1} = 15$.

(d) The true lower limit (X_l) of the modal class is noted from Table 3.9. Thus, $X_l = 59.5$.

$$d_1 = f_m - f_{m-1} = 28 - 14 = 14; \quad d_2 = f_m - f_{m+1} = 28 - 15 = 13.$$

$$M_o = X_l + i \times \frac{d_1}{d_1 + d_2} = 59.5 + 3 \times \frac{14}{14 + 13} = 61.1 \text{ kg.}$$

GLOSSARY

central tendencies : statistics of location such as mean, geometric mean and median which are specific scores near the middle of the frequency distribution, around which all the scores of the sample are distributed.

central values, measures of : same as central tendencies.

deciles : scores, below each of which there is a given number of one-tenths of all the scores of the sample.

fractiles : scores, below each of which lies a given fraction of all the scores, i.e., of the frequency distribution, of a sample.

mean : a measure of central value, worked out as the arithmetic average of the scores of a sample.

mean, geometric : a measure of central value, worked out as the n th root of the product of all the n number of scores of the sample.

mean, weighted : a mean of a number of samples, taken together, worked out from the sum of the products of the means of those samples and their respective sample sizes.

median : a measure of central value, below which lies 50% of all the scores, i.e., the lower half of the frequency distribution, in a sample.

mode : a measure of central value, being the score which has the highest frequency of occurrence in the sample and thus corresponds to the peak of the distribution.

percentiles : fractiles, below each of which lies a given percentage of the total number of scores in the sample.

percentile rank : a rank given to a score, on a scale of 100, among all the scores of a sample.

quantiles : same as fractiles.

quartiles : fractiles, below each of which lies a given number of quarters of the frequency distribution, i.e., of the total number of scores, in a sample.

sum of squares : the sum of the squared differences between the scores of a sample and the sample mean.

4. STATISTICS OF DISPERSION

Variability of a variable is the tendency of dispersion of its scores around the measures of central values such as the mean, median or mode of a sample. Statistics of dispersion serve as measures of variability of scores and give a quantitative idea about their frequency distributions.

4.1 MEASURES OF DISPERSION

These are such *descriptive statistics* (page 12) as describe the property of a sample with respect to the variability of the scores of a given variable in that sample. They measure and express numerically the deviations of the scores of a sample from a given central value like mean or median and thereby indicate the spread or scatter of the scores around that central value. They belong to the following two classes.

Absolute measures of dispersion :

These are worked out directly from the raw scores of a variable and are expressed in the same unit as that of those scores. They are widely used in many statistical work and include range, standard deviation, variance, quartile deviation and central moments. But (i) they cannot be used in comparing the dispersions or variabilities of the scores of more than one variable expressed in different units; for example, the standard deviations (*SD*) of body heights (cm) and body weights (kg) in a sample cannot be used for comparing their variabilities as these *SDs* are expressed in cm and kg respectively. Moreover, (ii) these absolute measures are not suitable for comparing the variabilities of two sets of scores, expressed in the same unit but having widely divergent central values like the mean ; for example, *SD* is not suitable for comparing the variabilities of femur lengths of giraffes and rats, both in cm units, because the mean and consequently the *SD* of the variable in giraffes far

surpass those of rats. (iii) These absolute measures are unsuitable for comparing different sets of data to find their differences in precision.

Relative measures of dispersion :

These are worked out, not directly from the raw scores of a variable, but from some absolute measures of dispersion and the corresponding measures of central values. Each relative measure is derived from a ratio of an absolute measure like *SD* and a central value like mean, is expressed as a percentage of the latter, and is consequently free from the unit of the corresponding raw scores of the variable. (i) This makes the relative measures very suitable for comparing the variabilities of two sets of scores given in different units. (ii) Having been expressed as percentages of central values, they are also preferable in comparing two sets of scores given in the same unit but diverging very widely from each other in their central values. (iii) Moreover, they can also compare the precision in different sets of data. Relative measures include coefficient of variation and coefficient of quartile deviation.

4.2 RANGE

Range is the interval from the highest to the lowest score of a measurement variable in a sample. Thus, all the scores of a sample fall within the range. In computations using the range, its value is taken as follows.

$$\text{Range} = [(\text{highest score}) - (\text{lowest score})] + 1.$$

It is the simplest absolute measure of dispersion, and bears the same unit as that of the raw scores. If the lowest and the highest fasting blood sugar scores are 60 and 114 mg dL⁻¹ respectively in a sample, the range is (114 - 60) + 1, or 55 mg; the range of IQ scores amounts to (105 - 80) + 1, or 26, in a sample with the minimum and maximum scores of respectively 80 and 105.

Range is used in working out quantitative frequency distributions and in statistical quality control. But it has many limitations : (i) The range cannot be found out for incomplete frequency distributions with *open class intervals* because the lowest and/or the highest scores of the sample are not available there. (ii) It indicates only two extreme scores of the sample with no information about the magnitudes and frequencies of intermediate scores. (iii) It gives no indication about the form of the distribution of scores — whether symmetric or skewed, whether unimodal or multimodal, whether mesokurtic, leptokurtic or platykurtic. (iv) It is very unstable and varies greatly with the sample size — the larger the sample, the greater is the chance of including more extreme scores and so, the wider is the range. (v) Even when all other scores are close to each other, a single extreme score may increase the range disproportionately.

4.3 MEAN DEVIATION

As the algebraic sum of the deviations of all scores of a sample from the mean, viz., $\Sigma(X - \bar{X})$, always amounts to 0, a measure of absolute dispersion cannot be straightway worked out as the arithmetic average of those deviations. So, mean deviation (MD) is sometimes computed as an absolute measure, using the sum of *absolute deviations* of the scores from the mean, disregarding the algebraic signs of those deviations.

$$MD = \frac{\Sigma |X - \bar{X}|}{n},$$

where the bars indicate the ignoring of the algebraic signs. MD bears the same unit as that of the raw scores. But (i) it measures the variations of scores in magnitude only, not in direction; so, (ii) it is unsuitable for many further statistical work.

4.4 STANDARD DEVIATION

Standard deviation (SD) is the positive square root of the mean of squared deviations of all the

scores from their mean. It is, in other words, the positive square root of variance or of the second moment about the mean (vide § 4.5 and 4.6). It is a very useful absolute measure of dispersion and is in the same unit as the original scores. The standard deviations of a sample and of a population are denoted by s and σ respectively. For a sample,

$$s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n}},$$

where X represents the individual scores, \bar{X} is the sample mean, $(X - \bar{X})$ or x is the deviation of a score from \bar{X} , and n is the total frequency or sample size; the sum of the squared deviations, viz., $\Sigma(X - \bar{X})^2$, used in this computation, is called the *sum of squares* (SS). For a population with size N and mean μ ,

$$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}.$$

Some properties of SD are as follows.

(a) Because SD takes into consideration all the scores of a sample, it changes with the change of even a single score.

(b) Addition or subtraction of a constant number to or from each individual score leaves the SD unaltered, but the multiplication or division of each score by a constant number produces an identical change in the SD.

(c) If all scores have an identical value in a sample, the SD amounts to 0.

(d) The higher the SD, the wider is the dispersion of scores around the mean.

(e) In different samples from the same population, SDs differ far less than do the other absolute measures of dispersion.

(f) If the scores of a variable Y are the linear functions of the scores of another variable X , the plotting of the Y scores against the X scores of the respective individuals produces a straight line with

a as its vertical intercept and b as its slope. Thus, for the X and Y scores of each individual,

$$Y = a + bX.$$

In such case, if s_Y and s_X are the SDs of the respective variables and the bars on the two sides of b indicate that the latter is taken as positive irrespective of its actual algebraic sign,

$$s_Y = |b| s_X.$$

(g) In a small sample ($n < 30$), extreme scores at the two ends of the frequency distribution may escape inclusion due to their less frequent occurrence in the population. Because SD depends on all the scores, the exclusion of many extreme scores from the sample tends to lower the SD of a small sample much below the population σ . The SD of a small sample thus suffers from an undesirable downward bias.

(h) The composite SD of a number of groups can be computed from their individual group sizes (n_i), the group SDs (s_i), and the deviations of respective group means (\bar{X}_i) from the grand mean (\bar{X}) of all the groups.

$$s = \sqrt{\frac{\sum n_i s_i^2 + \sum n_i (\bar{X}_i - \bar{X})^2}{\sum n_i}}.$$

Computation from simple ungrouped series

1. For large samples :

SD may be computed by any of the following formulae from the simple series of ungrouped scores (X) if the total frequency (n) of the sample is large (> 30).

Example 4.4.1.

Compute the SD of the following body weights (kg) of 16 men :

55, 60, 62, 58, 57, 61, 59, 60, 61, 62, 67, 48, 54, 65, 63, 52.

Solution :

(i) The data are arranged in Table 4.1 and the mean \bar{X} is computed using the sum of the scores.

$$\sum X = 944; \quad n = 16; \quad \therefore \bar{X} = \frac{\sum X}{n} = \frac{944}{16} = 59.0 \text{ kg.}$$

$$(i) \quad s = \sqrt{\frac{\sum (X - \bar{X})^2}{n}};$$

$$(ii) \quad s = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2};$$

$$(iii) \quad s = \sqrt{\frac{n \sum X^2 - (\sum X)^2}{n^2}};$$

$$(iv) \quad s = \sqrt{\frac{\sum X^2 - n \bar{X}^2}{n}}.$$

2. Unbiased SD for small samples :

The downward bias of SD of a small sample, due to the reason cited in the paragraph (g) above, may be compensated by using its degrees of freedom ($n - 1$) instead of n in its computation. Such an SD is called an *unbiased SD*. (For degrees of freedom, see § 5.5.)

$$(i) \quad s = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}};$$

$$(ii) \quad s = \sqrt{\frac{n \sum X^2 - (\sum X)^2}{n(n - 1)}};$$

$$(iii) \quad s = \sqrt{\frac{\sum X^2 - n \bar{X}^2}{n - 1}}.$$

Because even a large sample is smaller than the population, the SD of even such a sample, when computed with n as the denominator, suffers from some downward bias, though small enough to be neglected often. Hence, it is preferable to compute the unbiased SD for even large samples, using $(n - 1)$ instead of n as the denominator.

(ii) The deviation of each score from the mean, viz., $(X - \bar{X})$, is worked out, recorded with its algebraic sign, and squared (Table 4.1).

(iii) The sum of all the $(X - \bar{X})^2$ values is used in working out the unbiased SD (s).

$$s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}} = \sqrt{\frac{360}{16-1}} = 4.90 \text{ kg.}$$

Table 4.1. Table for computing mean and SD of body weights.

Serial No.	Body weight (X)	$X - \bar{X}$	$(X - \bar{X})^2$	Serial No.	Body weight (X)	$X - \bar{X}$	$(X - \bar{X})^2$
1	55	-4	16	Total brought forward	533		40
2	60	+1	1	10	62	+3	9
3	62	+3	9	11	67	+8	64
4	58	-1	1	12	48	-11	121
5	57	-2	4	13	54	-5	25
6	61	+2	4	14	65	+6	36
7	59	0	0	15	63	+4	16
8	60	+1	1	16	52	-7	49
9	61	+2	4				
Total carried forward	533		40	Total	944 (ΣX)		360 ($\Sigma(X - \bar{X})^2$)

Alternative solution :

(i) After entering the data in Table 4.2, each X score is squared and the squared scores are also entered in the table. Both X and X^2 scores are totalled to give ΣX and ΣX^2 , respectively.

(ii) The unbiased SD (s) is worked out using ΣX and ΣX^2 .

$$s = \sqrt{\frac{n\Sigma X^2 - (\Sigma X)^2}{n(n-1)}} = \sqrt{\frac{16 \times 56056 - 944^2}{16(16-1)}} = 4.90 \text{ kg.}$$

Table 4.2. Table for computing SD of body weights using raw scores.

Serial No.	Body weight (X)	X^2	Serial No.	Body weight (X)	X^2
1	55	3025	Total brought forward	533	31605
2	60	3600	10	62	3844
3	62	3844	11	67	4489
4	58	3364	12	48	2304
5	57	3249	13	54	2916
6	61	3721	14	65	4225
7	59	3481	15	63	3969
8	60	3600	16	52	2704
9	61	3721			
Total carried forward	533	31605	Total	944 (ΣX)	56056 (ΣX^2)

Example 4.4.2.

Compute the *SD* of the following memory test scores of 20 High School students :

9, 10, 12, 15, 9, 11, 16, 10, 13, 9, 12, 10, 14, 13, 15, 16, 13, 10, 12, 14.

Solution :

(i) After entering the data in Table 4.3, the mean \bar{X} is computed using the sum of the scores.

$$\Sigma X = 243 ; \quad n = 20 ; \quad \therefore \bar{X} = \frac{\Sigma X}{n} = \frac{243}{20} = 12.15.$$

Table 4.3. Table for computing mean and *SD* of memory test scores.

Individual	Test score (X)	X ²	Individual	Test score (X)	X ²
1	9	81	Total brought forward	126	1502
2	10	100	12	10	100
3	12	144	13	14	196
4	15	225	14	13	169
5	9	81	15	15	225
6	11	121	16	16	256
7	16	256	17	13	169
8	10	100	18	10	100
9	13	169	19	12	144
10	9	81	20	14	196
11	12	144			
Total carried forward	126	1502	Total	243 (ΣX)	3057 (ΣX^2)

(ii) Each *X* score is squared and the squared scores are totalled to give ΣX^2 which is used in computing the unbiased *SD*.

$$s = \sqrt{\frac{\Sigma X^2 - n\bar{X}^2}{n-1}} = \sqrt{\frac{3057 - 20 \times 12.15^2}{20-1}} = 2.346.$$

(iii) Alternatively :

$$s = \sqrt{\frac{n\Sigma X^2 - (\Sigma X)^2}{n(n-1)}} = \sqrt{\frac{20 \times 3057 - 243^2}{20(20-1)}} = 2.346.$$

Example 4.4.3.

Find the unbiased *SD* of the following winglength scores (mm) of a sample of cockroaches :

35, 36, 26, 28, 44, 30, 22, 33, 27, 25, 40, 44, 35, 31, 29, 32.

Solution :

(a) After tabulating the scores in Table 4.4, each *X* score is squared and the squared scores (*X*²) are entered in the table. Both *X* and *X*² scores are totalled to give ΣX and ΣX^2 respectively, which are used in working out the unbiased *SD*.

$$s = \sqrt{\frac{n\sum X^2 - (\sum X)^2}{n(n-1)}} = \sqrt{\frac{16 \times 17331 - 517^2}{16(16-1)}} = 6.46 \text{ mm.}$$

Table 4.4. Table for computing SD of winglengths from raw scores.

Serial No.	Winglengths (X)	X ²	Serial No.	Winglengths (X)	X ²
1	35	1225	Total brought forward	281	9119
2	36	1296	10	25	625
3	26	676	11	40	1600
4	28	784	12	44	1936
5	44	1936	13	35	1225
6	30	900	14	31	961
7	22	484	15	29	841
8	33	1089	16	32	1024
9	27	729			
Total carried forward	281	9119	Total	517 (ΣX)	17331 (ΣX ²)

(b) Alternatively :

(i) After tabulating the X scores in Table 4.5, the mean \bar{X} is computed from their sum.

$$\Sigma X = 517; n = 16; \therefore \bar{X} = \frac{\Sigma X}{n} = \frac{517}{16} = 32.3 \text{ mm.}$$

(ii) The deviation of each score from the mean, viz., $X - \bar{X}$, is worked out, recorded with its algebraic sign and squared (Table 4.5.).(iii) The sum of the squared deviations, $\Sigma(X - \bar{X})^2$, is used in working out the unbiased SD.

$$s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}} = \sqrt{\frac{625.44}{16-1}} = 6.46 \text{ mm.}$$

Table 4.5. Table for computing the mean and SD of winglengths.

Serial No.	X	$X - \bar{X}$	$(X - \bar{X})^2$	Serial No.	X	$X - \bar{X}$	$(X - \bar{X})^2$
1	35	+ 2.7	7.29	Total brought forward	281		356.01
2	36	+ 3.7	13.69	10	25	- 7.3	53.29
3	26	- 6.3	39.69	11	40	+ 7.7	59.29
4	28	- 4.3	18.49	12	44	+ 11.7	136.89
5	44	+ 11.7	136.89	13	35	+ 2.7	7.29
6	30	- 2.3	5.29	14	31	- 1.3	1.69
7	22	- 10.3	106.09	15	29	- 3.3	10.89
8	33	+ 0.7	0.49	16	32	- 0.3	0.09
9	27	- 5.3	28.09				
Total carried forward	281		356.01	Total	517 (ΣX)		625.44 (Σ(X - \bar{X}) ²)

Computation from grouped data of frequency distributions

This method applies to data arranged into frequency distributions with either equal or unequal class intervals. Because all scores of a class interval are assumed to be identical with its midpoint (X_c),

$$\Sigma X = \Sigma fX_c; \quad \Sigma (X - \bar{X})^2 = \Sigma f(X_c - \bar{X})^2;$$

where f and X_c are respectively the frequencies and midpoints of the intervals.

(i) The X_c of each interval is first worked out.

$$X_c = \frac{1}{2}[(\text{upper limit}) + (\text{lower limit})]$$

(ii) The frequency (f) of each interval is multiplied by its X_c and these products (fX_c) for all the

intervals are totalled to give ΣfX_c .

(iii) The mean (\bar{X}) is then computed.

$$\bar{X} = \frac{\Sigma fX_c}{n}.$$

(iv) The deviation of each X_c from \bar{X} is worked out and squared to give $(X_c - \bar{X})^2$.

(v) Each $(X_c - \bar{X})^2$ is multiplied by the f of that interval and these products are totalled to give $\Sigma f(X_c - \bar{X})^2$.

(vi) The unbiased SD is then computed.

$$s = \sqrt{\frac{\Sigma f(X_c - \bar{X})^2}{n-1}}.$$

Example 4.4.4.

Compute the mean and SD of body heights (cm) in the following distribution :

Class intervals :	156-160	161-165	166-170	171-175	176-180
Frequencies :	4	14	25	11	6

Solution :

(i) After entering the data in Table 4.6, the midpoint X_c of each interval is worked out from the upper and lower limits of that interval. For example, for the interval 161-165,

$$X_c = \frac{165 + 161}{2} = 163 \text{ cm.}$$

(ii) The frequency (f) of each interval is multiplied by its X_c and these products (fX_c) are totalled for all the intervals to give ΣfX_c which is used in computing the mean.

$$\bar{X} = \frac{\Sigma fX_c}{n} = \frac{10085}{60} = 168.1 \text{ cm.}$$

(iii) The difference between each X_c and the mean is worked out and squared to get the $(X_c - \bar{X})^2$ of the relevant interval.

Table 4.6. Table for computation of mean and SD of heights.

Class intervals	X_c	f	fX_c	$X_c - \bar{X}$	$(X_c - \bar{X})^2$	$f(X_c - \bar{X})^2$
156-160	158	4	632	-10.1	102.01	408.04
161-165	163	14	2282	-5.1	26.01	364.14
166-170	168	25	4200	-0.1	0.01	0.25
171-175	173	11	1903	+4.9	24.01	264.11
176-180	178	6	1068	+9.9	98.01	588.06
Total		60 (n)	10085			1624.60

(iv) Each $(X_c - \bar{X})^2$ is multiplied by the frequency f of that interval and such products of all the intervals are totalled to give $\Sigma f(X_c - \bar{X})^2$ which is used in working out the unbiased SD .

$$s = \sqrt{\frac{\Sigma f(X_c - \bar{X})^2}{n-1}} = \sqrt{\frac{1624.60}{60-1}} = 5.25 \text{ cm.}$$

Example 4.4.5.

Compute the mean and unbiased SD of the following distribution of Ca^{2+} concentration scores (in mg per kg of extracellular fluid) in a sample of lobsters.

Class intervals	: 11.6-13.0	13.1-14.5	14.6-16.0	16.1-17.5	17.6-19.0
Frequencies	: 7	13	20	14	6

Solution :

(i) After entering the data in Table 4.7, the midpoint X_c of each interval is worked out from the limits of that interval. For example, for the interval 14.6-16.0,

$$X_c = \frac{1}{2}[(\text{lower limit}) + (\text{upper limit})] = \frac{1}{2}(14.6 + 16.0) = 15.3.$$

(ii) The frequency (f) of each interval is multiplied by its X_c and the sum of such products of all the intervals, viz., ΣfX_c , is used in working out the mean.

$$\bar{X} = \frac{\Sigma fX_c}{n} = \frac{916.5}{60} = 15.3 \text{ mg.}$$

(iii) The difference between each X_c and the mean is worked out and squared to get $(X_c - \bar{X})^2$ of the relevant interval.

(iv) Each $(X_c - \bar{X})^2$ is multiplied by the frequency f of that interval and such products of all the intervals are totalled to give $\Sigma f(X_c - \bar{X})^2$ which is used in working out the unbiased SD .

$$s = \sqrt{\frac{\Sigma f(X_c - \bar{X})^2}{n-1}} = \sqrt{\frac{177.75}{60-1}} = 1.74 \text{ mg.}$$

Table 4.7. Table for computing mean and SD of Ca^{2+} concentrations.

Class intervals	X_c	f	fX_c	$X_c - \bar{X}$	$(X_c - \bar{X})^2$	$f(X_c - \bar{X})^2$
11.6-13.0	12.3	7	86.1	-3	9	63
13.1-14.5	13.8	13	179.4	-1.5	2.25	29.25
14.6-16.0	15.3	20	306.0	0	0	0
16.1-17.5	16.8	14	235.2	+1.5	2.25	31.50
17.6-19.0	18.3	6	109.8	+3	9	54
Total		60 (n)	916.5			177.75

Computation from simple frequency table

Where n number of scores are arranged individually — without any grouping — in a simple frequency table, the unbiased SD may be

computed by using the frequency f of each score and the mean \bar{X} of the scores.

$$s = \sqrt{\frac{\Sigma f(X - \bar{X})^2}{n-1}}.$$

Example 4.4.6.

Arrange the following body heights (cm) in a simple frequency table and compute their *SD* :
180, 165, 170, 162, 176, 167, 180, 162, 165, 165, 170, 170.

Solution :

(i) After entering the individual scores (X) and their respective frequencies (f) in Table 4.8, the mean \bar{X} is computed from the sum of the products (fX) of the scores and their respective frequencies.

$$\bar{X} = \frac{\sum fX}{n} = \frac{2032}{12} = 169.3 \text{ cm.}$$

Table 4.8. Simple frequency table of height scores for computing *SD*.

Height scores (X)	f	fX	$X - \bar{X}$	$(X - \bar{X})^2$	$f(X - \bar{X})^2$
162	2	324	-7.3	53.29	106.58
165	3	495	-4.3	18.49	55.47
167	1	167	-2.3	5.29	5.29
170	3	510	+0.7	0.49	1.47
176	1	176	+6.7	44.89	44.89
180	2	360	+10.7	114.49	228.98
Total	12 (n)	2032			442.68

(ii) The difference between each score and \bar{X} is computed and squared to get $(X - \bar{X})^2$ which is multiplied by the frequency f of that score. Such products for all the scores are totalled to give $\sum f(X - \bar{X})^2$ which is used in working out the unbiased *SD*.

$$s = \sqrt{\frac{\sum f(X - \bar{X})^2}{n-1}} = \sqrt{\frac{442.68}{12-1}} = 6.34 \text{ cm.}$$

4.5 VARIANCE

Variance or mean square (s^2 or *MS*) is the arithmetic mean of the squared deviations of individual scores from the mean. It is a good absolute measure of dispersion, identical with the squared *SD* as well as the second central moment (page 61); this is why *SD* is known as the *root-mean-square* about the mean. Variance is in squared units like cm^2 and kg^2 . It is used in the *analysis of variances* (*anova tests*).

The sample variance is generally computed as follows for ungrouped scores :

$$s^2 = \frac{\sum (X - \bar{X})^2}{n-1}, \quad \text{or} \quad s^2 = \frac{n \sum X^2 - (\sum X)^2}{n(n-1)},$$

where $\sum (X - \bar{X})^2$ is the sum of squared differences of scores from the sample mean and is known as the *sum of squares*, and $\sum X^2$ is the sum of the squared scores.

For a frequency distribution grouped into regular class intervals, variance is computed as the squared *SD* by the following formula :

$$s^2 = \frac{\sum f(X_c - \bar{X})^2}{n-1},$$

where f and X_c are respectively the frequencies and midpoints of the intervals. (For comparison with *SD*, see pages 52 and 56).

The parametric or population variance has the symbol σ^2 .

Example 4.5.1.

Compute the variance and SD of the following housefly winglengths (mm) :

3.5, 4.8, 4.3, 3.4, 5.1, 4.2, 3.8, 4.5, 3.6, 5.0, 3.4, 4.4, 5.3, 3.7, 4.0, 3.3.

Solution :

(a) After entering the scores (X) in Table 4.9, each X score is squared and the squared scores (X^2) are also entered in the table. Both X and X^2 scores are totalled to give ΣX and ΣX^2 , respectively, which are used in working out the variance (s^2) and the unbiased SD.

$$s^2 = \frac{n \Sigma X^2 - (\Sigma X)^2}{n(n-1)} = \frac{16 \times 281.23 - 66.3^2}{16(16-1)} = 0.433 \text{ mm}^2; \quad s = \sqrt{s^2} = \sqrt{0.433} = 0.658 \text{ mm}.$$

Table 4.9. Table for computing variance of housefly winglengths from raw scores.

Serial No.	Winglengths (X)	X^2	Serial No.	Winglengths (X)	X^2
1	3.5	12.25	Total brought forward	37.2	156.64
2	4.8	23.04			
3	4.3	18.49			
4	3.4	11.56			
5	5.1	26.01	10	5.0	25.00
6	4.2	17.64	11	3.4	11.56
7	3.8	14.44	12	4.4	19.36
8	4.5	20.25	13	5.3	28.09
9	3.6	12.96	14	3.7	13.69
Total carried forward	37.2	156.64	15	4.0	16.00
			16	3.3	10.89
			Total	66.3 (ΣX)	281.23 (ΣX^2)

(b) Alternatively :

(i) The X scores are tabulated in Table 4.10 and the mean \bar{X} is computed from their sum ΣX .

$$\Sigma X = 66.3; \quad n = 16; \quad \therefore \bar{X} = \frac{\Sigma X}{n} = \frac{66.3}{16} = 4.14 \text{ mm}.$$

(ii) The difference of each score from the mean, viz., $X - \bar{X}$, is worked out, recorded in the table and squared to give the $(X - \bar{X})^2$ scores.

(iii) The sum of these squared differences, viz., $\Sigma(X - \bar{X})^2$, is used in working out the variance.

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n-1} = \frac{6.4996}{16-1} = 0.433 \text{ mm}^2; \quad s = \sqrt{s^2} = \sqrt{0.433} = 0.658 \text{ mm}.$$

Table 4.10. Table for computing variance of housefly winglengths from their sum of squares.

Serial No.	X	$X - \bar{X}$	$(X - \bar{X})^2$	Serial No.	X	$X - \bar{X}$	$(X - \bar{X})^2$
1	3.5	-0.64	0.4096	Total brought forward	37.2		2.8804
2	4.8	+0.66	0.4356	10	5.0	+0.86	0.7396
3	4.3	+0.16	0.0256	11	3.4	-0.74	0.5476
4	3.4	-0.74	0.5476	12	4.4	+0.26	0.0676
5	5.1	+0.96	0.9216	13	5.3	+1.16	1.3456
6	4.2	+0.06	0.0036	14	3.7	-0.44	0.1936
7	3.8	-0.34	0.1156	15	4.0	-0.14	0.0196
8	4.5	+0.36	0.1296	16	3.3	-0.84	0.7056
9	3.6	-0.54	0.2916				
Total carried forward	37.2		2.8804	Total	66.3		6.4996
					(ΣX)		$\Sigma(X - \bar{X})^2$

Example 4.5.2.

Compute the variance and *SD* of the body height (cm) distribution of Example 4.4.4.

Solution :

(i) After entering the data in Table 4.11, X_c of each interval is worked out from score limits of that interval. For example, for the interval 166-170.

$$X_c = \frac{1}{2}[(\text{upper limit}) + (\text{lower limit})] = \frac{170 + 166}{2} = 168 \text{ cm.}$$

(ii) The frequency f of each interval is multiplied by its X_c and these products (fX_c) are totalled for all intervals to give ΣfX_c which is used in computing the mean.

$$\bar{X} = \frac{\Sigma fX_c}{n} = \frac{10085}{60} = 168.1 \text{ cm.}$$

Table 4.11. Table for computation of variance from grouped data of body heights.

Class interval	X_c	f	fX_c	$X_c - \bar{X}$	$(X_c - \bar{X})^2$	$f(X_c - \bar{X})^2$
156-160	158	4	632	-10.1	102.01	408.04
161-165	163	14	2282	-5.1	26.01	364.14
166-170	168	25	4200	-0.1	0.01	0.25
171-175	173	11	1903	+4.9	24.01	264.11
176-180	178	6	1068	+9.9	98.01	588.06
Total		60 (n)	10085			1624.60

(iii) The difference between each X_c and the mean is calculated and squared to give the $(X_c - \bar{X})^2$ of that interval.

(iv) Each $(X_c - \bar{X})^2$ is multiplied by the f of that interval and these products for all the intervals are totalled to give $\Sigma f(X_c - \bar{X})^2$ which is used in working out the variance.

$$s^2 = \frac{\Sigma f(X_c - \bar{X})^2}{n-1} = \frac{1624.60}{60-1} = 27.54 \text{ cm}^2; \quad s = \sqrt{s^2} = \sqrt{27.54} = 5.25 \text{ cm.}$$

4.6 CENTRAL MOMENTS

A *central moment about the mean* is the arithmetic average of the deviations of individual scores from the mean, each such deviation raised to a given power. Central moments are absolute measures of dispersion.

The *first central moment* (m_1) about the mean is computed from the sum of deviations, each raised to the power 1. It amounts to 0 for both symmetric or asymmetric distributions.

$$m_1 = \frac{\sum(X - \bar{X})}{n} = 0.$$

The *second central moment* (m_2) about the mean is the arithmetic average of the squared deviations of scores from the mean, and is identical with variance (s^2). It is expressed in squared units like cm^2 and kg^2 . Its square root is the *SD*.

$$m_2 = \frac{\sum(X - \bar{X})^2}{n} = s^2.$$

The *third and fourth central moments* (m_3 and m_4) are computed as follows, and are expressed in units raised to the powers 3 and 4 respectively.

$$m_3 = \frac{\sum(X - \bar{X})^3}{n}; \quad m_4 = \frac{\sum(X - \bar{X})^4}{n}$$

The third and higher central moments of the odd order amount to 0 in symmetric distributions, but possess positive or negative values according as the distribution is positively or negatively skewed. So, m_3 and higher central moments of the odd order serve as *measures of skewness* or asymmetry. The central moments of the even order, viz., m_2 and m_4 , are used in measuring the peakedness or *kurtosis* of a distribution.

4.7 QUARTILE DEVIATION

The *interquartile range* of a frequency distribution extends from the first quartile (Q_1 or P_{25}) to the third quartile (Q_3 or P_{75}). Thus, it contains the middle 50% of the scores of a

distribution. *Quartile deviation* (Q) or the *semi-interquartile range* constitutes half of this middle 50% range of scores. It is an absolute measure of dispersion, expressed in the same unit as the scores.

$$Q = \frac{Q_3 - Q_1}{2} = \frac{P_{75} - P_{25}}{2}.$$

Some of its *properties* are as follows.

(a) Because the lowest 25% and the highest 25% of the scores lie beyond the interquartile range from which Q is computed, the latter is independent of the scores at the two tails of the distribution. Thus, Q is unaffected by extreme scores at either tail of the distribution and can be computed even for incomplete distributions with open class intervals.

(b) Q is not affected by scores other than Q_3 and Q_1 . Thus, it gives no idea about the distribution and variability of other scores either within or beyond the interquartile range. This is a serious deficiency of Q .

(c) *Kurtosis* or peakedness of a frequency distribution is proportional to Q . The smaller the Q , the greater is the concentration of scores at the middle of the distribution and the more is the distribution *leptokurtic* with a high peak and a narrow body. A large Q shows a long interquartile range owing to a wider dispersal of scores of the middle order, making the distribution *platykurtic* with a low peak and a broad body.

(d) In unimodal and bilaterally symmetric distributions like the *normal distribution*, Q_2 or *Mdn* coincides with the mean and the mode at the centre of the distribution. In such cases, Q_3 and Q_1 are equidistant from Q_2 . Q then contains exactly 25% of the total scores on either side of the *Mdn* and equals 0.6745σ . In other words, the range $X \pm Q$ includes 50% of the scores in symmetric unimodal distribution.

(e) In asymmetric distributions, one tail of the distribution is skewed or more drawn out than the other. Q_3 and Q_1 are no longer equidistant from

Q_2 or Mdn . So, the midpoint of the interquartile range gets displaced towards the skewed tail. On two sides of Mdn , Q now covers unequal percentages of the total scores. Q is used in measuring skewness and kurtosis of a distribution.

Example 4.7.1.

Compute the quartile deviation of the following achievement test scores in a group of students.

Class intervals	:	81-90	91-100	101-110	111-120	121-130	131-140
Frequencies	:	7	12	19	24	14	4

Solution :

(i) After entering the data in Table 4.12, the true lower limit (X_l) for each interval and the cumulative frequency (cf) upto its X_u are computed. For example, for the 4th interval 101-110,

$$X_l = \frac{1}{2}[(\text{lower score limit of the interval}) + (\text{upper score limit of the next lower interval})]$$

$$= \frac{1}{2}(111 + 110) = 110.5.$$

$$cf = f_1 + f_2 + f_3 + f_4 = 7 + 12 + 19 + 24 = 62.$$

(ii) The size i of the intervals is obtained by subtracting the lower score limit of any interval from that of the next higher one. Using the lower score limits of 101-110 and 111-120,

$$i = 111 - 101 = 10.$$

Table 4.12. Cumulative frequencies of achievement test scores. (cf upto the respective X_u scores)

Class intervals	X_l	f	cf
131-140	130.5	4	80 (n)
121-130	120.5	14	76
111-120	110.5	24	62
101-110	100.5	19	38
91-100	90.5	12	19
81-90	80.5	7	7

(iii) For computing Q_1 or P_{25} , the proportion p of total cases to be counted of starting from the lowest interval, amounts to 0.25.

$$p = 0.25; \quad n = 80; \quad \therefore pn = 0.25 \times 80 = 20.$$

On counting off 20 scores starting from the lowest interval, the interval 101-110 is reached, in which P_{25} lies. The true lower limit (X_l) and the frequency (f_p) of that interval, and the cumulative frequency (cf_l) upto its X_l are noted from Table 4.12:

$$X_l = 100.5; \quad f_p = 19; \quad cf_l = 19.$$

$$\therefore P_{25} \text{ or } Q_1 = X_l + i \times \frac{pn - cf_l}{f_p} = 100.5 + 10 \times \frac{20 - 19}{19} = 101.0.$$

(iv) For computing Q_3 or P_{75} ,

$$p = 0.75; \quad n = 80; \quad \therefore pn = 0.75 \times 80 = 60.$$

On counting off 60 scores starting from the lowest interval, the interval 111-120 is reached in which P_{75} lies.

$$X_l = 110.5; \quad f_p = 24; \quad cf_l = 38.$$

$$\therefore P_{75} \text{ or } Q_3 = X_l + i \times \frac{pn - cf_l}{f_p} = 110.5 + 10 \times \frac{60 - 38}{24} = 119.7.$$

(v) The quartile deviation is then computed.

$$Q = \frac{Q_3 - Q_1}{2} = \frac{119.7 - 101.0}{2} = 9.35.$$

Example 4.7.2.

Compute the quartile deviation of the following distribution of cockroach winglength scores (mm).

Class intervals :	20-22	23-25	26-28	29-31	32-34
Frequencies :	6	14	20	12	8

Solution :

(i) After entering the data in Table 4.13, the lower limit (X_l) for each interval and the cumulative frequency (cf) upto its X_u are worked out. For example, for the 3rd interval 26-28,

$$X_l = \frac{1}{2}[(\text{lower score limit of the interval}) + (\text{upper score limit of the next lower interval})]$$

$$= \frac{1}{2}(26 + 25) = 25.5.$$

$$cf = f_1 + f_2 + f_3 = 6 + 14 + 20 = 40.$$

(ii) The interval size (i) is obtained by subtracting the lower score limit of any interval from that of the next higher interval. Using the lower limits of 26-28 and 29-31,

$$i = 29 - 26 = 3.$$

(iii) For computing Q_1 or P_{25} , the proportion p of total cases to be counted off, starting from the lowest interval, amounts to 15.

$$p = 0.25; \quad n = 60; \quad \therefore pn = 0.25 \times 60 = 15.$$

On counting off 15 scores starting from the lowest interval, the interval 23-25 is reached, in which P_{25} lies. The true lower limit (X_l), the cumulative frequency (cf_l) upto that X_l and the frequency (f_p) in that interval are noted from Table 4.13 and used in working out Q_1 or P_{25} .

$$X_l = 22.5; \quad cf_l = 6; \quad f_p = 14; \quad i = 3;$$

$$\therefore Q_1 \text{ or } P_{25} = X_l + i \times \frac{pn - cf_l}{f_p} = 22.5 + 3 \times \frac{15 - 6}{14} = 24.43 \text{ mm.}$$

Table 4.13. Cumulative frequencies of winglength scores. (cf upto the respective X_u scores)

Class intervals	X_l	f	cf
20-22	19.5	6	6
23-25	22.5	14	20
26-28	25.5	20	40
29-31	28.5	12	52
32-34	31.5	8	60 (n)

(iv) For computing Q_3 or P_{75} , similarly,

$$p = 0.75; \quad n = 60; \quad \therefore pn = 0.75 \times 60 = 45.$$

On counting off 45 scores starting from the lowest interval, the interval 29-31 is reached, in which P_{75} lies.

$$X_i = 28.5; \quad cf_i = 40; \quad f_p = 12; \quad i = 3;$$

$$\therefore Q_3 \text{ or } P_{75} = X_i + i \times \frac{pn - cf_i}{f_p} = 28.5 + 3 \times \frac{45 - 40}{12} = 29.75 \text{ mm.}$$

(v) Q_1 and Q_3 are then used in computing the quartile deviation (Q).

$$Q = \frac{Q_3 - Q_1}{2} = \frac{29.75 - 24.43}{2} = 2.66 \text{ mm.}$$

4.8 COEFFICIENT OF VARIATION

Coefficient of variation (CV) is a relative measure of dispersion, obtained by expressing the SD as a percentage of the mean.

$$CV = \frac{s}{\bar{X}} \times 100.$$

Expressed as a percentage, CV is independent of any particular unit, and suitable in comparing variabilities of two variables measured in different units. Thus, it can be used (i) in comparing variabilities of physicochemical variables

expressed in ratio scales (with true 0), such as serum cholesterol scores (mg) and arterial BP (mm Hg). (ii) It can also compare variabilities of two sets of scores in the same unit but with widely divergent means, e.g., femur lengths (cm) of giraffes and mice.

CV is not suitable (i) for psychological and educational data in interval scales (with no true 0 point) because no ratio can be computed with SD and mean in the latter scale, (ii) nor where the mean is close to 0 because CV may approach infinity there.

Example 4.8.1.

Find the relative variability of the following serum cholesterol scores (X) and systolic arterial pressure scores (Y) in a sample of humans.

$$\begin{array}{lll} \text{Serum cholesterol} & : & \bar{X} = 164.6 \text{ mg dL}^{-1}; \quad s_X = 18.86 \text{ mg.} \\ \text{Systolic pressure} & : & \bar{Y} = 128.6 \text{ mm Hg}; \quad s_Y = 13.74 \text{ mm.} \end{array}$$

Solution :

$$(i) \text{ For serum cholesterol, } CV_X = \frac{s_X}{\bar{X}} \times 100 = \frac{18.86}{164.6} \times 100 = 11.46.$$

$$(ii) \text{ For systolic pressure, } CV_Y = \frac{s_Y}{\bar{Y}} \times 100 = \frac{13.74}{128.6} \times 100 = 10.68.$$

$$(iii) \frac{CV_X}{CV_Y} = \frac{11.46}{10.68} = 1.07.$$

So, variability of cholesterol scores is 1.07 times that of pressure scores.

4.9 COEFFICIENT OF QUARTILE DEVIATION

$$CQD = \frac{Q}{Mdn} \times 100.$$

It is also a relative measure of dispersion. It is obtained by expressing the quartile deviation (Q) as a percentage of the median (Mdn or Q_2).

Being independent of the units of raw scores, it can be used to compare variabilities of two sets of scores in different units.

Example 4.9.1.

Compare the variabilities of body weight and body height scores of a sample, using the following data.

Body weight	:	$Q_1 = 58.2 \text{ kg};$	$Q_2 = 61.0 \text{ kg};$	$Q_3 = 63.7 \text{ kg}.$
Body height	:	$Q_1 = 153.4 \text{ cm};$	$Q_2 = 166.2 \text{ cm};$	$Q_3 = 176.2 \text{ cm}.$

Solution :

(i) For body weight :

$$Q_w = \frac{Q_3 - Q_1}{2} = \frac{63.7 - 58.2}{2} = 2.75 \text{ kg}.$$

$$CQD_w = \frac{Q_w}{Mdn} \times 100 = \frac{2.75 \text{ kg}}{61.0 \text{ kg}} \times 100 = 4.51.$$

(ii) For body height :

$$Q_h = \frac{Q_3 - Q_1}{2} = \frac{176.2 - 153.4}{2} = 11.4 \text{ cm}.$$

$$CQD_h = \frac{Q_h}{Mdn} \times 100 = \frac{11.4 \text{ cm}}{166.2 \text{ cm}} \times 100 = 6.86.$$

As the CQD is higher in case of height, variability of height scores is much higher than that of weight in the sample.

4.10 COEFFICIENT OF DISPERSION

This is the ratio of variance and mean of a sample.

$$CD = \frac{s^2}{\bar{X}}$$

Thus, it is a relative measure of dispersion, but bears the same unit as the scores.

Because CD bears the unit of the raw scores, it is not suitable in comparing the variabilities of two sets of scores expressed in two different units. However, it can be used for comparing variabilities of two sets of scores which have widely divergent means, but are given in the same unit. CD amounts to 1.00 in a distribution where events occur at

random and independent of each other. But CD is less than 1 in *repulsed distributions* where the occurrence of one event reduces the probability of occurrence of a second similar event and consequently, the centre and the two tails of the distribution carry respectively higher and lower frequencies of scores than in a distribution having random occurrences of events. On the contrary, CD exceeds 1 in *clumped distributions* where the occurrence of one event increases the probability of occurrence of a second similar event and consequently, the centre and the tails of the distribution carry respectively lower and higher frequencies than in a random distribution of events.

GLOSSARY

- central moments** : absolute measures of dispersion worked out as the arithmetic averages of deviations of scores from their mean, each such deviation having been raised to a given power.
- coefficient of dispersion** : a relative measure of dispersion given by the ratio between the variance and the mean of a sample.
- coefficient of quartile deviation** : a relative measure of dispersion which is the quartile deviation expressed as a percentage of the median.
- coefficient of variation** : a relative measure of dispersion obtained by expressing the standard deviation as a percentage of the mean.
- mean deviation** : an absolute measure of dispersion given by the arithmetic average of the absolute differences of the individual scores of a sample from their mean, disregarding the algebraic signs of the differences.
- measures of dispersion, absolute** : statistics of dispersion such as standard deviation, variance and quartile deviation, which are computed directly from raw scores of a sample and bear the same units as those of the raw scores.
- measures of dispersion, relative** : statistics of dispersion such as the coefficients of variation and quartile deviation, which are obtained by expressing an absolute measure of dispersion as the percentage or proportion of a central value like the mean.
- quartile deviation** : an absolute measure of dispersion given by half of the difference between the third and the first quartiles.
- range** : an absolute measure of dispersion which is the interval extending from the lowest score to the highest score of a variable in a sample.
- standard deviation** : an absolute measure of dispersion which is the square root of the mean of squared differences of the scores of a sample from the sample mean.
- statistics of dispersion** : statistics like standard deviation, variance and coefficient of variation, which measure and express in numerical units the dispersion of scores of a variable around central values like the mean and the median in a sample.
- sum of squares** : the sum of squared differences of the scores of a variable from a central value like the mean.
- variance** : an absolute measure of dispersion given by the mean of squared differences of the scores of a sample from their mean.

5. SAMPLING STATISTICS

The statistic of a variable varies from sample to sample drawn from the same population, and forms a frequency distribution around the corresponding population parameter. Sampling statistics such as standard errors serve as the measures of deviation of the statistics of a sample from the respective population parameters. A sampling statistic is also used in working out a confidence interval in which the corresponding parameter has a specified probability of falling. Sampling statistics also help in drawing generalized inferences for the entire population from the findings in one or more samples.

5.1 SAMPLING ERRORS

Each sample consists of only a limited number of individuals or cases drawn from a vast population. The cases and their scores vary from sample to sample ; even if the same individuals constitute successive samples, their scores are not identical in successive measurements due to their temporal variations. Consequently, any statistic of a sample frequently differs from similar statistics of other samples from the same population. Although the statistics of a few samples may coincide with the population parameter, the statistics of many more samples differ from the parameter. *Sampling error* of a statistic is the difference between a statistic and the corresponding parameter. Its cause lies in the chance factors associated with the random sampling of only a limited number of individuals, in exclusion of others, from the vast population. Sampling error (e) may be negative or positive according as the parameter exceeds the statistic or falls short of the latter ; thus, the sampling error of a sample mean would be : $e = \bar{X} - \mu$. For example, if the finite population of all the participants in an Olympiad has a mean vital

capacity (μ) of 5.8 litres and a random sample of the participants has a mean (\bar{X}) of 6.4 litres, the sampling error of \bar{X} is given by :

$$e = \bar{X} - \mu = 6.4 - 5.8 = + 0.6 \text{ litre.}$$

Similarly, if the mean IQ on the Stanford Binet scale amounts to 120 for the entire finite population of IIT-JEE candidates, and to 116 for a small sample of those candidates, the sampling error of the mean of this sample is given by :

$$e = \bar{X} - \mu = 116 - 120 = - 4.$$

Sampling errors may be estimated from the variations of the relevant statistic in samples drawn from the given population.

5.2 SAMPLING DISTRIBUTIONS

If many large samples of identical size n are drawn by random sampling from the same population and a particular statistic (say, the mean) is worked out from the scores of each sample, the computed values of the statistic are distributed in a frequency distribution, called an *experimental sampling distribution* of that statistic, around the corresponding parameter. A similar frequency distribution of a statistic of different samples from a population of a known nature can also be constructed theoretically using the laws of probability ; this gives a *theoretical sampling distribution* of that statistic. Any sampling distribution is due to the varying sampling errors of the relevant statistic of different samples from the corresponding population parameter.

The standard deviation of a sampling distribution of a statistic is a measure of the dispersion of the latter around the corresponding population parameter, and is called the *standard error* (SE) of that statistic. For example, the means of samples drawn from a population form a

sampling distribution of means and the standard deviation of this distribution is called the *SE of the mean* ($s_{\bar{X}}$).

Sampling distributions can also be framed for the differences between the statistics of different samples drawn from the same or different populations. Standard deviation of such a sampling distribution is a measure of dispersion of differences between the statistics of each pair of samples around the difference between the corresponding population parameters; it is known as the *SE of the difference* between the relevant statistics. For example, differences between the means of every pair of samples from the same or two separate populations form a *sampling distribution of differences between the means*, and the standard deviation of this distribution is the *SE of the difference between the means* ($s_{\bar{X}_1 - \bar{X}_2}$).

5.3 STANDARD ERRORS

Standard error (*SE*) of a statistic is a measure of its sampling error which is the deviation of that statistic from the corresponding parameter. It is the standard deviation of the sampling distribution of the relevant statistic – the standard deviation of the theoretical sampling distribution is the *parametric or true SE* (e.g., $\sigma_{\bar{X}}$) while the standard deviation of an experimental sampling distribution is an *estimate of the true SE* and is computed from the observed sample data.

SE is computed for many sample statistics such as mean, standard deviation, proportion and correlation coefficient. It is used (i) in measuring the variability of a statistic between different samples of a population, (ii) in testing the significance of the relevant statistic, and (iii) in computing a *confidence interval* within which the population parameter has a specified probability of falling. (iv) The standard error of the statistics of two samples, drawn from the same or different populations, may be used in computing the *SE of the difference* between such statistics.

The *SE of the difference* between the sample statistics of a particular type is the standard deviation of the sampling distribution of such differences around the difference between the parameters of the populations from which the samples have been drawn. It is a measure of the variability of such differences and is used in testing experimental hypotheses by finding the significance of difference between the statistics of two samples exposed to different levels of experimental treatments.

Standard error of the mean

Means of samples from a population form a *sampling distribution of means* around the parametric mean μ of that population; the standard deviation of this distribution is the *SE of the means*. For a theoretical sampling distribution of means (§ 5.2), the standard deviation of the distribution is the *true or population SE* ($\sigma_{\bar{X}}$) of the means; the standard deviation of an experimental sampling distribution (§ 5.2) is an *estimate of $\sigma_{\bar{X}}$* and is represented by $s_{\bar{X}}$.

$s_{\bar{X}}$ is a measure of the deviation of sample means from the population mean, and an index of the sampling error of means. It is used (i) in finding whether or not \bar{X} is a dependable estimate of μ , (ii) in working out a confidence interval in which μ has a specified probability of falling, and (iii) in computing the *SE of the difference* between the means of different samples for testing experimental hypotheses (chapter 7).

$s_{\bar{X}}$ can be conveniently estimated even by using a single sample — its value is inversely proportional to the square root of the size of the sample used.

(a) For a sample of size n , drawn from a *finite population* of size N by *simple random sampling without replacement*:

$$(i) \ s_{\bar{X}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}},$$

where s is the unbiased SD given by :

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} ; \text{ but}$$

$$(ii) s_{\bar{X}} = \frac{s}{\sqrt{n-1}} \sqrt{\frac{N-n}{N-1}}, \text{ where}$$

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n}}$$

(b) For a sample drawn from an infinite population ($N = \infty$) by simple random sampling without replacement :

$$(i) s_{\bar{X}} = \frac{s}{\sqrt{n}}, \text{ where } s = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}, \text{ but}$$

$$(ii) s_{\bar{X}} = \frac{s}{\sqrt{n-1}}, \text{ where } s = \sqrt{\frac{\sum (X - \bar{X})^2}{n}}$$

(c) In case the sample has been drawn by simple random sampling with replacement, irrespective of whether the population is finite or infinite,

$$(i) s_{\bar{X}} = \frac{s}{\sqrt{n}}, \text{ where } s = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} ;$$

$$(ii) s_{\bar{X}} = \frac{s}{\sqrt{n-1}}, \text{ where } s = \sqrt{\frac{\sum (X - \bar{X})^2}{n}}$$

(d) If the sample is drawn by stratified random sampling from a population with the means of its different strata differing negligibly, $s_{\bar{X}}$ is computed by the same formulae as used in case of simple random sampling.

(e) If the sample is drawn by stratified random sampling from a population having substantial differences between the means of its different strata,

$$s_{\bar{X}} = \sqrt{\frac{s^2 - s_{st}^2}{N}},$$

where s is the unbiased SD of the entire sample, N is the total sample size and s_{st} is the SD of the sample strata.

Where X_1, X_2 , etc., are the raw scores of different strata, \bar{X}_1, \bar{X}_2 , etc., are the respective stratum means, n_1, n_2 , etc., are the respective stratum sizes, and \bar{X} is the grand mean, s_{st}^2 is worked out as follows for this computation :

$$s_{st}^2 = \frac{n_1(\bar{X}_1 - \bar{X})^2 + \dots + n_k(\bar{X}_k - \bar{X})^2}{N}, \text{ or}$$

$$s_{st}^2 = \frac{1}{N} \left[\frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \dots + \frac{(\sum X_k)^2}{n_k} \right] - \left[\frac{\sum X_1 + \sum X_2 + \dots + \sum X_k}{N} \right]^2$$

Example 5.3.1.

Compute the SE of the mean using the following frequency distribution of body heights (cm).

Class intervals	:	156-160	161-165	166-170	171-175	176-180
Frequencies	:	4	14	25	11	6

Solution :

The frequency distribution is entered in Table 5.1. The midpoint (X_c) of each interval is computed using the two score limits of that interval. For example, for the interval 166-170 (Table 5.1),

$$X_c = \frac{170 + 166}{2} = 168.$$

$$\bar{X} = \frac{\sum fX_c}{n} = \frac{10085}{60} = 168.1 \text{ cm.}$$

$$s = \sqrt{\frac{\sum f(X_c - \bar{X})^2}{n-1}} = \sqrt{\frac{1624.60}{60-1}} = 5.25 \text{ cm.}$$

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{5.25}{\sqrt{60}} = 0.678 \text{ cm.}$$

Table 5.1. Frequency distribution of body heights for computation of SE.

Class intervals	X_c	f	fX_c	$X_c - \bar{X}$	$(X_c - \bar{X})^2$	$f(X_c - \bar{X})^2$
156-160	158	4	632	-10.1	102.01	408.04
161-165	163	14	2282	-5.1	26.01	364.14
166-170	168	25	4200	-0.1	0.01	0.25
171-175	173	11	1903	+4.9	24.01	264.11
176-180	178	6	1068	+9.9	98.01	588.06
Total		60 (n)	10085			1624.60

Example 5.3.2.

Compute the SE of the mean hemoglobin concentration (g dl^{-1}) of the following data in a stratified sample from a population divided on the basis of sex.

Men : 14.2, 14.6, 14.8, 14.0, 13.7, 14.8, 14.6, 13.8, 14.0, 14.9, 15.0.
 Women : 12.4, 12.8, 12.6, 13.1, 12.5, 12.9, 13.4, 14.2, 12.9, 12.2.

Solution :

The data are arranged in Table 5.2. The stratum sizes in the sample are as follows :

Men : $n_1 = 11$. Women : $n_2 = 10$.

(a) The stratum means \bar{X}_1 and \bar{X}_2 and the grand mean \bar{X} are computed.

$$\bar{X}_1 = \frac{\sum X_1}{n_1} = \frac{158.4}{11} = 14.4 \text{ g; } \bar{X}_2 = \frac{\sum X_2}{n_2} = \frac{129.0}{10} = 12.9 \text{ g;}$$

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2} = \frac{11 \times 14.4 + 10 \times 12.9}{11 + 10} = 13.7 \text{ g.}$$

(b) Deviations of scores of each stratum from \bar{X} are computed, squared and totalled. These sums of squared deviations are used in computing the variance s^2 of the entire sample.

$$s^2 = \frac{\sum (X_1 - \bar{X})^2 + \sum (X_2 - \bar{X})^2}{n_1 + n_2 - 1} = \frac{2.22 + 2.98}{(11 + 10) - 1} = 0.26 \text{ g}^2.$$

(c) The squared deviation of each stratum mean from \bar{X} is then used in computing the variance s_{st}^2 of strata.

$$s_{st}^2 = \frac{n_1(X_1 - \bar{X})^2 + n_2(X_2 - \bar{X})^2}{n_1 + n_2} = \frac{11(14.4 - 13.7)^2 + 10(12.9 - 13.7)^2}{11 + 10} = 0.56 \text{ g}^2.$$

Alternatively :

$$s_{st}^2 = \frac{1}{N} \left[\frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} \right] - \left[\frac{\sum X_1 + \sum X_2}{N} \right]^2 = \frac{1}{21} \left[\frac{158.4^2}{11} + \frac{129^2}{10} \right] - \left[\frac{158.4 + 129}{21} \right]^2 = 0.56 \text{ g}^2.$$

(d) The SE ($s_{\bar{X}}$) of the mean is then computed using s^2 and s_{st}^2 .

$$s_{\bar{X}} = \sqrt{\frac{s^2 + s_{st}^2}{N}} = \sqrt{\frac{0.26 + 0.56}{11 + 10}} = 0.198 \text{ g.}$$

Table 5.2. Hemoglobin data for computing SE of a stratified sample.

Men			Women		
X_1	$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	X_2	$X_2 - \bar{X}_2$	$(X_2 - \bar{X}_2)^2$
14.2	-0.2	0.04	12.4	-0.5	0.25
14.6	+0.2	0.04	12.8	-0.1	0.01
14.8	+0.4	0.16	12.6	-0.3	0.09
14.0	-0.4	0.16	13.1	+0.2	0.04
13.7	-0.7	0.49	12.5	-0.4	0.16
14.8	+0.4	0.16	12.9	0.0	0.00
14.6	+0.2	0.04	13.4	+0.5	0.25
13.8	-0.6	0.36	14.2	+1.3	1.69
14.0	-0.4	0.16	12.9	0.0	0.00
14.9	+0.5	0.25	12.2	-0.7	0.49
15.0	+0.6	0.36			
Σ 158.4		2.22	129.0		2.98

Example 5.3.3.

Compute the SE of the mean, variance and the unbiased SD of the following distribution of housefly winglength scores ($\text{mm} \times 10^{-1}$).

Class intervals :	34-37	38-41	42-45	46-49	50-53
Frequencies :	4	8	15	7	6

Solution :

After entering the frequency distribution in Table 5.3, the midpoint X_c of each interval is worked out using the two score limits of that interval. For example, for the interval 42-45,

$$X_c = \frac{1}{2}[(\text{upper score limit}) + (\text{lower score limit})] = \frac{1}{2}(45 + 42) = 43.5.$$

Table 5.3. Frequency distribution of winglength scores for computation of SE.

Class intervals	f	X_c	fX_c	$X_c - \bar{X}$	$(X_c - \bar{X})^2$	$f(X_c - \bar{X})^2$
34-37	4	35.5	142.0	- 8.3	68.89	275.56
38-41	8	39.5	316.0	- 4.3	18.49	147.92
42-45	15	43.5	652.5	- 0.3	0.09	1.35
46-49	7	47.5	332.5	+ 3.7	13.69	95.83
50-53	6	51.5	309.0	+ 7.7	59.29	355.74
Total	40 (n)		1752.0			876.40

$$\bar{X} = \frac{\sum fX_c}{n} = \frac{1752.0}{40} = 43.8 \times 10^{-1} \text{ mm.}$$

$$s^2 = \frac{\sum f(X_c - \bar{X})^2}{n-1} = \frac{876.40}{40-1} = 22.47 \times 10^{-2} \text{ mm.}$$

$$s = \sqrt{\frac{\sum f(X_c - \bar{X})^2}{n-1}} = \sqrt{\frac{876.40}{40-1}} = 4.74 \times 10^{-1} \text{ mm.}$$

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{4.74}{\sqrt{40}} = 0.749 \times 10^{-1} \text{ mm.}$$

Standard error of proportions

Where the population is divided into two classes with respect to a variable, and p and q are the proportions of cases of the respective classes in a sample of size n , the SE of the proportion p is worked out as follows.

$$s_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{pq}{n}}.$$

Standard error of difference between means

If pairs of samples are drawn repeatedly, the two samples of each pair coming from two different populations, and the difference between the sample means of every pair is worked out, these differences ($\bar{X}_1 - \bar{X}_2$) between the sample means would form a sampling distribution around the difference ($\mu_1 - \mu_2$) between the respective population means. For samples from the same population instead of two different ones, the

sampling distribution of the difference ($\bar{X}_1 - \bar{X}_2$) between their means would have a mean of 0, because μ_1 equals μ_2 in that case.

The standard deviation of any such sampling distribution of ($\bar{X}_1 - \bar{X}_2$) is the SE of the difference between the sample means, $s_{\bar{X}_1 - \bar{X}_2}$. It is a measure of the deviations of differences between sample means ($\bar{X}_1 - \bar{X}_2$) from a mean difference (μ_D) which equals either ($\mu_1 - \mu_2$) or 0, according as the samples have come from different populations or from an identical one. $s_{\bar{X}_1 - \bar{X}_2}$ is used in testing the significance of difference between two sample means in an experiment, and is computed from the standard errors ($s_{\bar{X}_1}$ and $s_{\bar{X}_2}$) of the respective sample means and the respective sample sizes,

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{(s_{\bar{X}_1})^2 + (s_{\bar{X}_2})^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Example 5.3.4.

The heights of 60 men and 50 women had unbiased SDs of 36.7 and 40.6 cm respectively. Compute the SE of the difference between mean heights of men and women.

Solution :

$$(a) \text{ For men : } n_1 = 60 ; s_1 = 36.7 \text{ cm ; } \therefore s_{\bar{X}_1} = \frac{s_1}{\sqrt{n_1}} = \frac{36.7}{\sqrt{60}} = 4.74 \text{ cm.}$$

$$(b) \text{ For women : } n_2 = 50 ; s_2 = 40.6 \text{ cm ; } \therefore s_{\bar{X}_2} = \frac{s_2}{\sqrt{n_2}} = \frac{40.6}{\sqrt{50}} = 5.74 \text{ cm.}$$

$$(c) s_{\bar{X}_1 - \bar{X}_2} = \sqrt{(s_{\bar{X}_1})^2 + (s_{\bar{X}_2})^2} = \sqrt{(4.74)^2 + (5.74)^2} = 7.44 \text{ cm.}$$

Example 5.3.5.

The unbiased standard deviations of the annual milk yields (in L $\times 10^2$) of 49 two-year-old Tharpakhna cows and 64 two-year-old Jersey cows were found to be 3.93 and 5.58, respectively. Work out the SE of difference between the mean milk yields of the two breeds.

Solution :

$$(a) \text{ For Tharpakhna cows : } n_1 = 49 ; s_1 = 3.93 ; \therefore s_{\bar{X}_1} = \frac{s_1}{\sqrt{n_1}} = \frac{3.93}{\sqrt{49}} = 0.561.$$

$$(b) \text{ For Jersey cows : } n_2 = 64 ; s_2 = 5.58 ; \therefore s_{\bar{X}_2} = \frac{s_2}{\sqrt{n_2}} = \frac{5.58}{\sqrt{64}} = 0.698.$$

(c) Using the respective SE scores of the two means,

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{(s_{\bar{X}_1})^2 + (s_{\bar{X}_2})^2} = \sqrt{0.561^2 + 0.698^2} = 0.896.$$

Standard error of difference between proportions respectively,

This is the standard deviation of the sampling distribution of differences between the proportions of cases of a given class in two samples. This $s_{p_1 - p_2}$ is a measure of the deviation of such a difference between the respective population proportions. It is used in testing the significance of a difference between two samples with respect to the proportions of a given type of cases.

(a) For samples from two different populations:

Where p_1 and p_2 are the proportions of a given class in two samples, s_{p_1} and s_{p_2} are their respective SEs, and q_1 and q_2 equal $(1 - p_1)$ and $(1 - p_2)$

$$s_{p_1} = \sqrt{\frac{p_1 q_1}{n_1}} ; \quad s_{p_2} = \sqrt{\frac{p_2 q_2}{n_2}} ;$$

$$s_{p_1 - p_2} = \sqrt{s_{p_1}^2 + s_{p_2}^2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}.$$

(b) For samples from the same population :
The common proportion p and q for both samples are first computed for use in working out $s_{p_1 - p_2}$.

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} ; \quad q = 1 - p ;$$

$$s_{p_1 - p_2} = \sqrt{\frac{pq}{n_1} + \frac{pq}{n_2}}.$$

Example 5.3.6.

Proportions of diabetics were found to be 0.32 in a sample of 144 men and 0.22 in a sample of 121 women. Work out the *SE* of the difference in the proportion of diabetics between the sexes.

Solution :

Where p_1 and p_2 are the proportions of diabetics, and q_1 and q_2 those of non-diabetics in men and women, respectively,

(a) for men :

$$n_1 = 144 ; \quad p_1 = 0.32 ; \quad q_1 = 1 - p_1 = 1 - 0.32 = 0.68 ;$$

$$s_{p_1} = \sqrt{\frac{p_1 q_1}{n_1}} = \sqrt{\frac{0.32 \times 0.68}{144}} = 0.039 ;$$

(b) for women

$$n_2 = 121 ; \quad p_2 = 0.22 ; \quad q_2 = 1 - p_2 = 1 - 0.22 = 0.78 ;$$

$$s_{p_2} = \sqrt{\frac{p_2 q_2}{n_2}} = \sqrt{\frac{0.22 \times 0.78}{121}} = 0.038.$$

(c) using the values of s_{p_1} and s_{p_2} ,

$$s_{p_1 - p_2} = \sqrt{s_{p_1}^2 + s_{p_2}^2} = \sqrt{0.039^2 + 0.038^2} = 0.054.$$

5.4 STANDARD SCORES

Raw scores are often transformed into *standard scores* for gaining a standard meaning, a common reference value and a comparability. Such transformations may be linear or nonlinear.

Linear transformations may change the zero point of the scale and the unit of measurement of the scores. These alter the mean and *SD* of the raw scores, but the differences between the transformed scores correspond closely to those between the respective raw scores in relative magnitude. So, the original shape, skewness and kurtosis of the distribution of raw scores remain unaltered in that of the transformed scores. Where X_s is the transformed score, \bar{X}_s and s_s are respectively the mean and *SD* of the transformed scores, and \bar{X} and s respectively those of the raw scores, any raw score (X) may be linearly

transformed by putting the desired values to \bar{X}_s and s_s in the following equation :

$$X_s = \bar{X}_s + (X - \bar{X}) \frac{s_s}{s}.$$

The simplest standard score is the linearly transformed *z score*, *standard deviate* or *relative deviate*. It is a measure of deviation of a raw score from the mean in terms of the standard deviation. On putting 0 as \bar{X}_s , and 1 as s_s , in the above-mentioned equation, the transformed score (X_s) is the *z score*. Thus

$$z = \frac{X - \bar{X}}{s}, \quad \text{or, } z = \frac{X - \mu}{\sigma}.$$

Thus, the *z score* expresses the deviation of a given raw score from the mean as so many times the *SD*. Irrespective of the original unit of raw

scores, z scores are expressed in standard deviation units (σ units). Evidently, the z score for the mean (μ) of the distribution amounts to 0. It is positive for any score higher than the mean, and negative for any score lower than the mean. Thus, a raw score which is higher than the mean by an amount equal to 1.65 times the SD , has the z score of $+1.65\sigma$; a raw score that is lower than the mean by 1.96 times the SD , has the z score of -1.96σ .

Being derived by linear transformation, the frequency distribution of z scores of a sample has the same shape and form as the original raw scores — only the unit is changed to σ units and the zero point of the scale has also been altered. Evidently, the mean of the z score distribution amounts to 0 and is identical with the z score for μ while the SD of the distribution amounts to 1. So, z scores are eminently suitable for comparing the scores of the same or different distribution(s) expressed in different units or with widely different means and SD s.

Just like the raw scores, the sample mean may also be similarly transformed into the z score, using the SE of the mean ($s_{\bar{X}}$).

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}, \quad \text{or, } z = \frac{\bar{X} - \mu}{s_{\bar{X}}}.$$

The z score so computed expresses the deviation of the sample mean from the population mean in terms of the SE . Evidently, the z score for μ , viz., $z = (\mu - \mu)/\sigma_{\bar{X}}$, amounts to 0, around which the z scores of the sample means form a sampling distribution identical in shape and form with the sampling distribution of the corresponding sample means — it may be recalled that z scores are linear transformations of raw

scores and consequently have the same form of sampling distribution as those raw scores.

The z score is also computed as the standard score to express the deviation of the difference ($\bar{X}_1 - \bar{X}_2$) between two sample means from the difference ($\mu_1 - \mu_2$) between the parametric means of the two populations, from which the samples have been drawn. Where $s_{\bar{X}_1 - \bar{X}_2}$ is the estimated SE of such differences between sample means,

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}}.$$

Where both the samples come from the same population,

$$\mu_1 - \mu_2 = 0; \quad \therefore z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}.$$

The z score for the difference between two sample means finds application in testing the significance of such a difference observed in an experiment. However, because the unit of z scores is 1σ which is relatively large, the computed z score often has a value with decimal.

Nonlinear transformations change not only the mean and SD , but also the shape, skewness and kurtosis of the original raw score distribution. Nonlinear transformations such as logarithmic, reciprocal, square-root and arc-sine transformations are often tried for converting a non-normal distribution of raw scores into a normal distribution of transformed scores; in psychology, nonlinear transformations like percentile ranks, T scores, C scores, stanines and mental age scores are frequently used. (See the chapters on *Analysis of variances* and *Psychological test construction* for details).

Example 5.4.1.

The means and SD s of winglengths in a sample of houseflies and in one of mosquitoes are as follows :

Houseflies :	Mean (\bar{X}_1) = 3.67 mm ;	$SD (s_1)$ = 0.181 mm.
Mosquitoes :	Mean (\bar{X}_2) = 2.12 mm ;	$SD (s_2)$ = 0.124 mm.

Compare the deviations of winglengths of a given housefly ($X_1 = 3.23$ mm) and a given mosquito ($X_2 = 2.23$ mm) from the respective sample means.

Solution :

$$\text{Housefly : } z_1 = \frac{X_1 - \bar{X}_1}{s_1} = \frac{3.23 - 3.67}{0.181} = -2.43\sigma$$

$$\text{Mosquito : } z_2 = \frac{X_2 - \bar{X}_2}{s_2} = \frac{2.23 - 2.12}{0.124} = +0.89\sigma$$

So, the winglength score of the given housefly is lower than the mean by -2.43σ units while that of the given mosquito is higher than its mean by $+0.89\sigma$.

5.5 DEGREES OF FREEDOM

The *degrees of freedom* (*df*) of a statistic amount to that number of scores of a variable in a sample, which can be changed freely in magnitude and sign without causing any alteration in the values of all such statistics that have to be used in its computation as the estimates of the respective parameters.

In computing a statistic, one or more of the already computed statistics may be used as estimates of the respective population parameters. Serving as an estimate of a parameter, each such pre-computed statistic used in the computation must remain unchanged during the operation; this causes the loss of freedom of any one of the scores of the sample to undergo any change, because free changes of all other scores of the sample would determine or fix the only remaining score if the pre-computed statistic has to remain unaltered. As the total number of scores constitutes the sample size n , the *df* is lowered from n by 1 for keeping each precomputed statistic fixed or unchanged. Hence, the *df* of a statistic is often given by the sample size n less the number m of the pre-

computed statistics used in its computation as estimates of parameters : $df = n - m$.

For example, the computation of *SD* of a sample needs the use of the sample mean \bar{X} as an estimate of the population mean μ ; consequently, s has the *df* of $(n - 1)$. But the computation of the pooled *SD* (\hat{s}) of two samples, having sizes n_1 and n_2 , involves the similar use of two sample means \bar{X}_1 and \bar{X}_2 as the estimates of the respective parametric means of the populations from which the samples have been drawn. So, the *df* of \hat{s} amounts to $(n_1 + n_2 - 2)$.

If the computation of a statistic uses two or more pre-computed statistics as the estimates of parameters, its *df* equals the sum of the *df* values of all those statistics. For example, the *SE* of the difference between two means, viz., $s_{\bar{X}_1 - \bar{X}_2}$, is computed using the *SDs* (s_1 and s_2) of the two samples; because these *SDs* have degrees of freedom given by $(n_1 - 1)$ and $(n_2 - 1)$ respectively, $s_{\bar{X}_1 - \bar{X}_2}$ as well as the *t* score computed from the latter has the *df* of $(n_1 + n_2 - 2)$.

GLOSSARY

degrees of freedom : for a statistic under consideration, that number of scores or cases of a sample which can be changed freely without affecting the values of all pre-computed statistics used in its computation as estimates of the respective parameters.

- linear transformation** : a transformation of raw scores of a variable into such scores as are free of the units of the raw scores, differ from the raw scores in mean and standard deviation, but retain the original shape of the distribution of the raw scores.
- nonlinear transformation** : a transformation of raw scores of a variable into such scores as are not only free of the units of the raw scores and differ from them in mean and SD , but also have a distribution different in form from that of the raw scores.
- sampling distribution** : a frequency distribution of a given statistic of different samples around the parameter of the population from which the samples have been drawn.
- sampling error** : the difference between a statistic of a sample and the corresponding parameter of the population from which the sample has been drawn.
- sampling statistic** : a statistic such as the standard error which goes beyond a single sample to estimate sampling errors and to make inferences.
- standard error** : a statistic estimating the differences between the statistics of samples and the corresponding parameter of the population from which the samples have been drawn.
- standard error of differences between means** : a statistic to estimate the deviations of differences between sample means from the difference between parametric means of the populations from which those samples have been drawn.
- standard error of differences between proportions** : a statistic to estimate the deviations of differences between sample proportions from the difference between proportions in the populations from which the samples have been drawn.
- standard error of mean** : a statistic to estimate the differences of sample means from the parametric mean of the population from which those samples have been drawn.
- standard error of proportions** : a statistic to estimate the differences between proportions of cases of a given class in samples from the relevant proportion in the population.
- standard score** : a linearly or nonlinearly transformed score, free of the unit of the raw score from which it has been worked out, and comparable with other similarly transformed scores for making inferences.
- z score** : a linearly transformed score in SD (σ) unit, having a normal distribution if the corresponding raw scores have a normal distribution.

6. PROBABILITY DISTRIBUTIONS

Experimental data have frequently to be either interpreted or used for predictions, depending theoretically on the laws of probability. Probability distributions are distributions of relative frequencies of events or cases, computed theoretically on the basis of mathematical models and laws of probability. They are used to find the probability of the data conforming to or violating the hypotheses being tested in experiments. Of the probability distributions frequently used, normal and t distributions are continuous distributions while binomial and Poisson distributions are discrete ones.

6.1 PROBABILITY

Probability (P) of the occurrence of an event is the limit approached by the relative frequency of that type of event in an infinitely large number of observations or trials. If, in a vast set of trials or observations numbering n , the occurrence of a given type of event is expected to attain a limiting frequency f_e , the probability of that event is given by :

$$P = \frac{f_e}{n}$$

But the observed frequency of the event may deviate widely from the expected limiting frequency if the total n number of events or trials is small. Nevertheless, its frequency approaches the expected limiting frequency with the increase in the number of trials or observations. The number of alternative events for each toss of a perfect coin amounts to two only, with either the head or the tail coming up. But if it is tossed only twice ($n = 2$), either the head or the tail may come up both the times in some cases while in other cases, both the head and the tail may come up once each. On the contrary, if an infinitely large number of tosses be performed, the number of

times the head (or the tail) comes up will approach half the total number of tosses because there are only two alternative types of events of equal weightage. So, the limiting relative frequency or probability of this event of a given side coming up is given by $\frac{1}{2}$ or 0.5. While choosing a single individual at random from a population having 60% males and 40% females, the probability of a male being chosen amounts to $\frac{60}{100}$ or 0.6. The probability of drawing a sickle cell anemia patient at random from a population with 5% sickle cell anemia cases would similarly come to $\frac{5}{100}$ or 0.05.

If the occurrence of an event is not influenced by the occurrence or non-occurrence of another event and vice versa, the two are *mutually independent events* ; thus, the second event has an unaltered probability of occurrence irrespective of whether the first event has occurred or not. Events that have an equal probability of occurrence are called *equally likely events* ; in a trial, any of such a set of events has the same chance of occurring. If two events influence and alter one another's probability of occurrence, they are *dependent events* ; dependent events give rise to either clumped or repulsed distributions (§ 4.10). If the occurrence of one event is prevented by the occurrence of another event and vice versa, they cannot occur together and are called *mutually exclusive events* ; the probability of the simultaneous occurrence of two exclusive events amounts to 0 : $P(1.2) = 0$. Such events, at least one of which is sure to occur in a trial, constitute a set of *exclusive events*.

Addition theorem :

This theorem states that the probability of occurrence of any one of k number of alternative and *mutually exclusive* events is given by the sum of the probabilities of their individual occurrences.

$$P(1 \text{ or } 2 \text{ or } \dots k) = P(1) + P(2) + \dots + P(k).$$

For example, if one rat has to be chosen at random from amongst 10 rats, the probability of a given rat No.1 getting chosen in one attempt is given by : $P(1) = 1/10 = 0.10$. The probability of a second rat No.2 getting chosen in one attempt also amounts to : $P(2) = 0.10$. The two choices are mutually exclusive as the choice of one nullifies the chance of the other being chosen by a single attempt. In such a case, the probability that either the rat No.1 or the rat No.2 will be drawn in one choice is given by :

$$P(1 \text{ or } 2) = P(1) + P(2) = 0.10 + 0.10 = 0.20.$$

Multiplication theorem :

This theorem states that the probability of combined (simultaneous or successive) occurrence of k number of *independent events* is given by the product of the probabilities of their separate individual occurrences.

$$P(1.2.3 \dots k) = P(1) \times P(2) \times P(3) \times \dots \times P(k).$$

For instance, the probability of drawing at random any given rat out of 10 rats by a single choice amounts to 0.10. Now, if three rats are chosen simultaneously or successively at random, replacing each chosen rat into the group before the next choice, the probability of choosing three particular rats, say, rat Nos.1, 3 and 7, is given by :

$$\begin{aligned} P(1.3.7.) &= P(1) \times P(3) \times P(7) \\ &= 0.10 \times 0.10 \times 0.10 \\ &= 0.001. \end{aligned}$$

6.2 PROBABILITY DISTRIBUTIONS

The *relative frequency* of scores or events in a class interval of a frequency distribution is obtained by dividing the frequency of that interval by the sample size n . If n is very large, this relative frequency (f/n) may be an estimate of the probability of scores or events occurring in that interval. Computation of probabilities of scores or events for all the class intervals of a frequency

distribution changes the latter into a *probability distribution* which is a distribution of probabilities of occurrence of scores, events or cases among the classes of the given variable (Table 6.1). A probability curve may be graphically plotted with the scores of the variable scaled along the X axis and their probabilities along the Y axis. In this way, a probability distribution may be computed *experimentally*, using the observed frequencies of scores or events in the data of a test or experiment.

Table 6.1. A probability distribution using observed frequencies.

Class intervals	f	$P = \frac{f}{n}$
60-64	18	0.075
65-69	24	0.100
70-74	42	0.175
75-79	78	0.325
80-84	30	0.125
85-89	36	0.150
90-94	12	0.050
Total	240 (n)	1.000

Theoretical probability distributions, on the contrary, are computed theoretically on the basis of specific mathematical models and laws of probability. They are used widely in predicting probabilities of events and in testing experimental hypotheses. Examples include the normal distribution computed on the basis of the Gaussian equation, Student's t distribution based on Gossett's equation, the probability distribution of rare events in terms of Poisson's equation, and the binomial distribution based on the binomial equation. In any probability distribution, probabilities are in a continuous scale with no real gaps between them. But the scale for the events, cases or scores of the variable under consideration may be either continuous or discontinuous. Thus, a probability distribution of scores of a discontinuous variable (e.g., cell counts, litter sizes, family sizes, pulse rates and respiratory

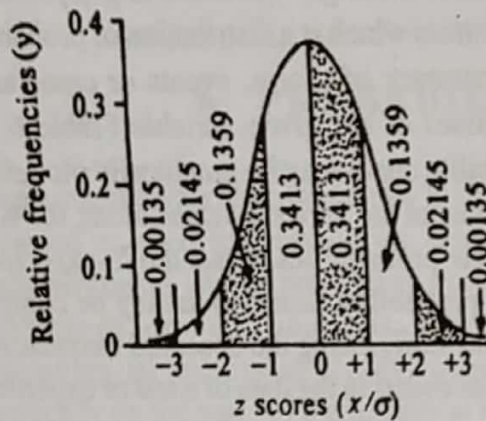


Fig. 6.1. Normal curve showing fractional areas between ordinates at different z scores. Note that the z score for μ is zero.

rates) or of cases in one of the discrete classes of a discontinuous variable (e.g., numbers of males or of mutant animals in samples) has real gaps in its scale for the scores or cases, and is called a *discrete probability distribution*; binomial and Poisson distributions are such discrete probability distributions. On the contrary, a *continuous probability distribution* like the normal or the t distribution is a distribution of probabilities of a continuous variable with no gap in its scale of scores, such as body weights, femur lengths, blood sugar or serum iron concentrations and anxiety test scores.

6.3 NORMAL DISTRIBUTION

Very often a specific form of a bilaterally symmetrical, unimodal, bell-shaped distribution, called the *normal distribution curve*, results on plotting the frequencies (f) of scores of a continuous measurement variable, observed in a very large sample, against the respective scores (X). A distribution of similar shape is obtained if the *relative frequencies* (f/n), obtained by dividing each observed frequency by the total frequency n , are plotted against the respective *standard scores* (z scores) computed from the raw X scores, because z scores are derived by a linear transformation of X scores (vide § 5.4); however, this distribution is called a *normal probability*

distribution because its Y ordinate gives the relative frequencies or probabilities, instead of the observed frequencies, of the respective z scores as also of the corresponding X scores (Fig. 6.1.).

Properties

1. The normal probability distribution gives the probable distribution of scores of a continuous measurement variable according to the laws of probability. It is thus a *continuous probability distribution* (§ 6.2).

2. Normal probability distributions can be *theoretically* computed, using the *Gaussian equation* to work out the probabilities of z scores. Where e is the base of the natural logarithm and amounts to 2.71828, π is the ratio between the circumference and the diameter of a circle and amounts to 3.14159, z is given by $(X - \mu)/\sigma$, and Y is the probability of the corresponding z score, the Gaussian equation formulated by Karl Friedrich Gauss (1777 – 1855) expresses Y as a function of z and so, of X .

$$Y = \frac{n}{\sigma\sqrt{2\pi}} e^{-(X-\mu)^2/2\sigma^2};$$

$$\text{or, } Y = \frac{n}{\sigma\sqrt{2\pi}} e^{-z^2/2}.$$

Imp.

Scaling Y along the ordinate and z along the abscissa, and plotting each computed Y against the corresponding z score, a normal curve is obtained with its peak at the z score for μ .

3. An infinite number of normal curves may be plotted depending on the sample size, the mean and the *SD*. But for reference in all cases, the *unit normal curve* is the standard form. It is computed taking the sample size (n), the *SD* (σ) and the length (i) of the class intervals of the distribution as 1.00 each. The Gaussian equation of the unit normal curve thus turns into :

$$Y = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

4. According to the *central theorem of probability*, a variable has an almost normal distribution if its scores depend on the independent effects of many other variables, acting at random and having no interaction among themselves. Many characteristics of living organisms have normal or near-normal distribution in the population, because their scores result largely from the effects of numerous genetic and environmental factors.

5. The z score for μ is the mean of the normal probability distribution and amounts to 0, because

$$z = \frac{\mu - \mu}{\sigma} = 0.$$

It has already been stated above that the *SD* of the unit normal curve is taken as 1.00.

6. The ordinate Y_o at the mean or centre of the normal curve is the highest of all its ordinates. For the unit normal curve, this ordinate is represented by y_o which measures 0.3989. The Y_o of any other normal curve ($n \neq 1$, $\sigma \neq 1$, $i \neq 1$) can be computed from y_o .

$$y_o = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} = \frac{1}{\sqrt{2 \times 3.14159}} e^0 = 0.3989;$$

$$Y_o = y_o \frac{in}{\sigma}; \quad \text{or, } Y_o = 0.3989 \frac{in}{s}.$$

7. The normal curve is *unimodal* and possesses a perfect *bilateral symmetry* around the single central peak. This makes the mean, median and mode coincide with the centre: $\mu = Mdn = M_o$.

8. Because of the bilateral symmetry, the first and third quartiles (Q_1 and Q_3) are equidistant from μ : $Q_3 = \mu + 0.6745\sigma$; $Q_1 = \mu - 0.6745\sigma$. Thus, the range of the normal distribution from -0.6745σ to $+0.6745\sigma$ covers the middle 0.5000 of the total area of the distribution (Table 6.2); the quartile deviation Q covers exactly 0.2500 of that area and amounts to 0.6745σ .

9. The normal distribution, being bilaterally symmetrical, is *free from skewness* — its

coefficient of skewness amounts to zero (§ 6.5).

10. The normal distribution is taken as a standard for the degree of peakedness or *kurtosis*. It is *mesokurtic* — its percentile coefficient of kurtosis is 0.263 and its moment coefficient is zero (§ 6.6).

11. The normal curve has *asymptotic tails*, progressively nearing the abscissa or X axis, but not reaching the latter except at infinite distances from the centre. So, the normal probability distribution extends from $-\infty$ to $+\infty$.

12. The total area of the normal curve represents the net frequency n . The latter is taken as 1 in case of the unit normal curve; so, the area of the latter is assumed as 1.0000. The ordinate Y_o at the mean bisects the area of the normal curve into two equal halves; so, each half of a unit normal curve has an area of 0.5000.

13. The fraction of the total frequency n , lying between two specified z scores, is given by the area between the ordinates at those z scores and is expressed as a fraction of 1.0000 which corresponds to the entire area under the normal curve (Fig. 6.2). Thus, the proportions of scores or cases lying between the mean μ ($z = 0$) and different multiples of σ are given by the fractional areas between the ordinates at μ and the z scores corresponding to the respective σ multiples.

14. Because of the bilateral symmetry of the unit normal curve, its fractional area between any two given z scores is *identical in both halves* of the curve. Thus, the fractional area between the z scores of +1 (i.e., $\mu + 1\sigma$) and +2 (i.e., $\mu + 2\sigma$) is identical with that between the z scores of -1 (i.e., $\mu - 1\sigma$) and -2 (i.e., $\mu - 2\sigma$), and amounts to 0.1359 (Fig. 6.1).

15. Because the area in each half of the unit normal curve is 0.5000, the fractional area in one tail beyond a given z score is given by the difference between 0.5000 and the fractional area from the mean μ ($z = 0$) to the given z . This gives

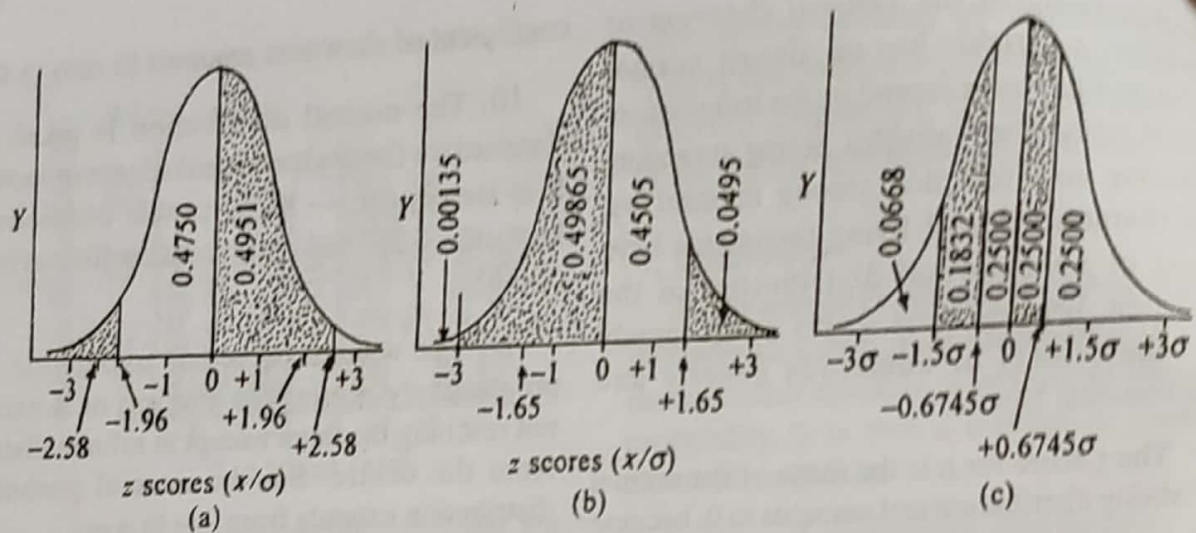


Fig. 6.2. Some fractional areas of the unit normal curve.

the probability P of cases having z scores equal to or beyond the given z in a single tail; thus, this fractional area in one tail beyond the given z gives the P of either (i) the cases with z scores higher than and equal to the given $+z$, or (ii) the cases with z scores equal to and lower than the given $-z$. So, for such *one-tail probabilities*,

$$P = 0.5000 - (\text{fractional area from the mean to the given } z).$$

A *one-tail critical z score* (z_{α}) is that z score beyond which lies a specified fractional area, also called the *one-tail level of significance* (α), in a single tail of the unit normal curve. Thus, the one-tail critical $z_{0.05}$ is the z score of 1.65, beyond which the single tail of the unit normal curve has a fractional area of 0.05 (Fig. 6.2b, Tables 6.2 and 7.1, § 7.3 and § 7.5).

16. Because the fractional areas in each half of the unit normal curve equal the corresponding ones in the other half, the total fractional area in the two tails beyond a given $\pm z$ score amounts to double the difference between 0.5000 and the fractional area from the mean ($z = 0$) to the given z in any one of the halves. This fractional area in two tails gives the probability P of the cases having z scores equal to or beyond the given z score in both halves of the normal curve. So, for such *two-tail probabilities*,

$$P = 2[0.5000 - (\text{fractional area from the mean to the given } z \text{ score})].$$

A *two-tail critical z score* (z_{α}) is that z score, beyond whose positive and negative values lies a specified total fractional area (α), also called the *two-tail level of significance*, in two tails of the

Table 6.2. Some fractional areas of the unit normal curve.

z scores in σ units	Intervals ($\mu \pm z$ score)	Fractional areas inside intervals	Fractional areas and probabilities beyond intervals	
			in one tail	in two tails
0.6745	$0 \pm 0.6745\sigma$	0.5000	0.2500	0.5000
1.00	$0 \pm 1\sigma$	0.6826	0.1587	0.3174
1.65	$0 \pm 1.65\sigma$	0.9010	0.0495	0.0990
1.96	$0 \pm 1.96\sigma$	0.9500	0.0250	0.0500
2.58	$0 \pm 2.58\sigma$	0.9902	0.0049	0.0098
3.00	$0 \pm 3\sigma$	0.9973	0.0014	0.0027

unit normal curve. For instance, the two-tail critical $z_{.05}$ is the z score of 1.96, beyond whose positive and negative values in the two tails lies a total fractional area of 0.05 (Tables 6.2 and 7.1, § 7.3 and § 7.5).

17. The normal curve shows upward convexity over the interval $\mu \pm 1\sigma$, i.e., between the z scores of +1 and -1, but becomes concave over both the tails beyond that interval. The central fractional area between these two z scores amounts to 0.6826 in the unit normal curve. Thus, 0.3413 of the total area lies between the ordinates at the mean and at the z score of either +1.00 or -1.00 (Fig. 6.1).

18. Means of random samples from a normally distributed population form a normal sampling distribution with the population mean μ at its centre (*sampling theory of means*).

19. Means of large samples from a population with a finite variance (σ^2) have an almost normal sampling distribution around μ , even if the variable is non-normally distributed in the population (*central limit theorem*).

Table A at the end of this book gives the ordinates (y) of a unit normal curve at different z (i.e., x/σ) scores as well as the fractional areas in one half of the curve from the mean μ ($z = 0$) to different z scores.

6.4 BEST-FITTING NORMAL DISTRIBUTION

It is that normal distribution which fits best with an observed distribution, and has the same mean (\bar{X}), the same SD and the same sample size

(n) as the latter. It can be computed as follows.

(a) The observed scores are arranged in a continuous frequency distribution with class intervals of equal size (i).

(b) The midpoint X_c of each class interval as well as \bar{X} and s of the sample is computed.

(c) Each X_c is transformed into a z score.

$$z = \frac{X_c - \bar{X}}{s}$$

(d) The unit normal curve table (Table A) is used to find the height y of the ordinate of the unit normal curve at each computed z score.

(e) Because i , n and s amount to 1.00 each in case of the unit normal curve, the ordinate Y of the best-fitting normal curve, corresponding to each recorded y of the unit normal curve, is then computed using the i , n and s of the given frequency distribution.

$$Y = y \times \frac{in}{s}$$

Each Y so computed gives the expected frequency (f_e) of the best-fitting normal distribution, for the class interval whose X_c corresponds to the z score for that Y . Thus, the computed values of Y constitute the distribution of f_e values for the best-fitting normal distribution (Table 6.4).

(f) Each Y score may be graphically plotted against the X_c of the corresponding class interval for drawing the best-fitting normal curve.

Example 6.4.1.

Compute the best-fitting normal distribution for the following frequency distribution of serum iron concentration ($\mu\text{g dL}^{-1}$) of 80 humans.

Class intervals :	100-109	110-119	120-129	130-139	140-149	150-159	160-169
Frequencies :	6	11	10	17	16	13	7

Solution :

(a) The data are entered in the first two columns of Table 6.3. The interval sizes (i) and the midpoints (X_c) of class intervals are worked out. For example, for the interval 120-129,

$$X_c = \frac{1}{2}[(\text{upper score limit}) + (\text{lower score limit})] = \frac{1}{2}(129 + 120) = 124.5.$$

$$i = (\text{lower score limit of an interval}) - (\text{lower score limit of next lower interval}) = 120 - 110 = 10.$$

Table 6.3. Table for computing mean and SD of serum iron data.

Class intervals	f	X_c	fX_c	$X_c - \bar{X}$	$(X_c - \bar{X})^2$	$f(X_c - \bar{X})^2$
100-109	6	104.5	627.0	- 31.6	998.56	5991.36
110-119	11	114.5	1259.5	- 21.6	466.56	5132.16
120-129	10	124.5	1245.0	- 11.6	134.56	1345.60
130-139	17	134.5	2286.5	- 1.6	2.56	43.52
140-149	16	144.5	2312.0	+ 8.4	70.56	1128.96
150-159	13	154.5	2008.5	+ 18.4	338.56	4401.28
160-169	7	164.5	1151.5	+ 28.4	806.56	5645.92
Σ	80 (n)		10890.0			23688.80

(b) Each X_c is multiplied by the corresponding f to give fX_c and the sum of the latter scores, viz., ΣfX_c , is used in working out \bar{X} .

$$\bar{X} = \frac{\Sigma fX_c}{n} = \frac{10890.0}{80} = 136.1 \mu\text{g}.$$

(c) The deviation of each X_c from \bar{X} is squared and the squared deviation is multiplied by the corresponding f ; the sum of these products is used in computing SD.

$$s = \sqrt{\frac{\Sigma f(X_c - \bar{X})^2}{n-1}} = \sqrt{\frac{23688.80}{80-1}} = 17.32 \mu\text{g}.$$

(d) The deviation of each X_c from \bar{X} is next transformed into z score which is entered in Table 6.4. For example, for the class interval 120-129,

$$X_c - \bar{X} = -11.6; \quad z = \frac{X_c - \bar{X}}{s} = \frac{-11.6}{17.32} = -0.67.$$

Table 6.4. Computation of the best-fitting normal distribution for serum iron data.

Class intervals	f	X_c	$X_c - \bar{X}$	z	y	Y or f_c
100-109	6	104.5	- 31.6	- 1.82	0.0761	3.5
110-119	11	114.5	- 21.6	- 1.25	0.1826	8.4
120-129	10	124.5	- 11.6	- 0.67	0.3187	14.7
130-139	17	134.5	- 1.6	- 0.09	0.3973	18.4
140-149	16	144.5	+ 8.4	+ 0.48	0.3555	16.4
150-159	13	154.5	+ 18.4	+ 1.06	0.2275	10.5
160-169	7	164.5	+ 28.4	+ 1.64	0.1040	4.8

(e) Neglecting the algebraic signs of the computed z scores, the height y of the ordinate at each z score is then recorded from the unit normal curve table (Table A). For example, for the z score of -0.67 , $y = 0.3187$.

(f) The height Y of the ordinate of the best-fitting normal distribution is computed for each z score by multiplying its y score with in/s . For example, for the class interval 120-129,

$$\frac{in}{s} = \frac{10 \times 80}{17.32} = 46.19; \quad Y = y \times \frac{in}{s} = 0.3187 \times 46.19 = 14.7.$$

The computed Y scores correspond to the expected frequencies (f_e) in the respective class intervals of the best-fitting normal distribution.

Example 6.4.2.

Work out the best-fitting normal distribution for the frequency distribution of anxiety test scores of 255 high-school girls presented in the first two columns of Table 6.5.

Solution :

(a) The midpoints (X_c) and the size (i) of the class intervals are worked out. For example, for the class interval 27-31,

$$X_c = \frac{1}{2}[(\text{upper score limit}) + (\text{lower score limit})] = \frac{1}{2}(31 + 27) = 29.$$

$$i = (\text{lower score limit of an interval}) - (\text{lower score limit of the next lower interval}) = 32 - 27 = 5.$$

Table 6.5. Table for computing mean and SD of anxiety score data.

Class intervals	f	X_c	fX_c	$X_c - \bar{X}$	$(X_c - \bar{X})^2$	$f(X_c - \bar{X})^2$
52-56	2	54	108	+ 29.6	876.16	1752.32
47-51	5	49	245	+ 24.6	605.16	3025.80
42-46	6	44	264	+ 19.6	384.16	2304.96
37-41	6	39	234	+ 14.6	213.16	1278.96
32-36	21	34	714	+ 9.6	92.16	1935.36
27-31	54	29	1566	+ 4.6	21.16	1142.64
22-26	63	24	1512	- 0.4	0.16	10.08
17-21	53	19	1007	- 5.4	29.16	1545.48
12-16	33	14	462	- 10.4	108.16	3569.28
7-11	11	9	99	- 15.4	237.16	2608.76
2-6	1	4	4	- 20.4	416.16	416.16
Σ	255 (n)		6215			19589.80

(b) Each X_c is multiplied by the corresponding f to give fX_c and the sum of the latter scores for all the intervals, viz., ΣfX_c , is used in working out \bar{X} .

$$\bar{X} = \frac{\Sigma fX_c}{n} = \frac{6215}{255} = 24.4.$$

(c) The difference between each X_c and \bar{X} is squared and the squared difference is multiplied by the corresponding f ; the sum of these products is used in computing SD .

$$s = \sqrt{\frac{\sum f(X_c - \bar{X})^2}{n-1}} = \sqrt{\frac{19589.80}{255-1}} = 8.78.$$

(d) The difference between each X_c and \bar{X} is next transformed into z score which is entered in Table 6.6. For example, for the class interval 27-31,

$$X_c - \bar{X} = 29 - 24.4 = +4.6; \quad z = \frac{X_c - \bar{X}}{s} = \frac{4.6}{8.78} = +0.52.$$

(e) Neglecting the algebraic signs of the computed z scores, the height y of the ordinate at each of them is then recorded from the unit normal curve table (Table A). For instance, for the z score of 0.52, $y = 0.3485$.

(f) The height Y of the ordinate of the best-fitting normal distribution is computed for each z score by multiplying its y score with in/s . For example, for the class interval 27-31,

$$i = 5; \quad n = 255; \quad s = 8.78; \quad \frac{in}{s} = \frac{5 \times 255}{8.78} = 145.2;$$

$$Y = y \times \frac{in}{s} = 0.3485 \times 145.2 = 50.6.$$

The computed Y scores correspond to the expected frequencies (f_e) in the respective class intervals of the best-fitting normal distribution.

Table 6.6. Computation of the best-fitting normal distribution for anxiety score data.

Class intervals	f	X_c	$X_c - \bar{X}$	z	y	Y or f_e
52-56	2	54	+ 29.6	+ 3.37	0.0014	0.2
47-51	5	49	+ 24.6	+ 2.80	0.0079	1.1
42-46	6	44	+ 19.6	+ 2.23	0.0332	4.8
37-41	6	39	+ 14.6	+ 1.66	0.1006	14.6
32-36	21	34	+ 9.6	+ 1.09	0.2203	32.0
27-31	54	29	+ 4.6	+ 0.52	0.3485	50.6
22-26	63	24	- 0.4	- 0.05	0.3984	57.8
17-21	53	19	- 5.4	- 0.62	0.3292	47.8
12-16	33	14	- 10.4	- 1.18	0.1989	28.9
7-11	11	9	- 15.4	- 1.75	0.0863	12.5
2-6	1	4	- 20.4	- 2.32	0.0270	3.9

6.5 SKEWNESS

In some frequency distributions, scores are concentrated at one end of the scale and are much fewer towards the other end. Such an asymmetric

distribution has its peak or mode towards the former end and a longer and more pointed tail at the other end. Such a distribution is called a *skewed distribution*. *Skewness* is the degree of asymmetry of the distribution.

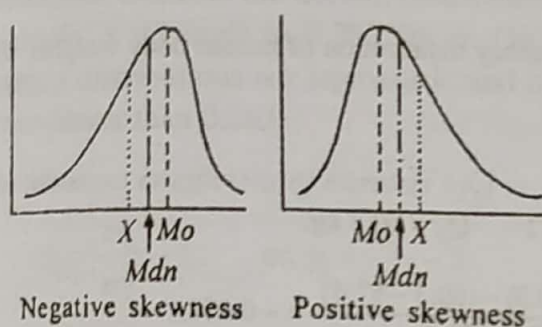


Fig. 6.3. Skewed distributions.

Properties of skewed distributions

(a) A skewed distribution cannot be bisected into two symmetrical halves, because one of its tails is longer and more tapering than the other.

(b) Skewness may be positive or negative according as the pointed longer tail rolls down to the high-value (right) end or the low-value (left) end of the scale, respectively (Fig. 6.3). The scores are more concentrated in the respective opposite halves of the distribution — agewise frequency distributions for atherosclerosis, cataract, malignancy and prostatic enlargement are negatively skewed as their incidences rise in old age.

(c) Mean, median and mode fail to coincide in an asymmetric distribution. Both median and mean are displaced from the mode towards the skewed tail, but the displacement of the mean considerably exceeds that of the median. So, $\bar{X} > Mdn > Mo$ in positively skewed distributions while $Mo > Mdn > \bar{X}$ in negatively skewed ones. Because the deflection of \bar{X} exceeds that of Mdn , the latter is more dependable of the two as a measure of central value in a skewed distribution.

(d) Unlike symmetrical distributions where all odd-order central moments possess a zero value, m_3 and higher odd-order central moments have either positive or negative values in skewed distributions, their signs and magnitudes indicating respectively the direction and the degree of skewness.

(e) Unlike symmetrical distributions where the first and third quartiles (Q_3 and Q_1) are equidistant from the second quartile (Q_2 or Mdn), Q_1 is displaced towards the skewed tail in an asymmetric distribution. Therefore, $(Q_3 - Q_2) > (Q_2 - Q_1)$ in positively skewed distributions, and $(Q_2 - Q_1) > (Q_3 - Q_2)$ in case of negative skewness.

Measures of skewness

(a) *Pearson's first coefficient of skewness* :

$$Sk = \frac{\bar{X} - Mo}{s}$$

(b) *Pearson's second coefficient of skewness* :

$$Sk = \frac{3(\bar{X} - Mdn)}{s}$$

The second coefficient is preferable to the first because of the frequent difficulty in estimating the mode of a distribution precisely. In symmetric distributions, $\bar{X} = Mdn = Mo$; so, both the Pearson coefficients amount to zero in such distributions.

(c) *Bowley's quartile coefficient of skewness* :

$$Sk = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

In any symmetric distribution, $Q_3 - Q_2 = Q_2 - Q_1$; so, the quartile coefficient amounts to zero in such distributions.

(d) *Moment coefficient of skewness (γ_1)* :

$$\gamma_1 = \frac{m_3}{s^3}$$

where the third moment m_3 about the mean is divided by the third power of SD ; because m_3 amounts to 0 for symmetric distributions, γ_1 amounts to 0 for such distributions.

All four coefficients are free from the unit of the variable. A positive or negative value indicates respectively a positive or negative skewness. The degree of skewness is given by the magnitude of the coefficient.

Example 6.5.1.

Compute the quartile coefficient of skewness for the frequency distribution of human body weights where Q_1 , Q_2 and Q_3 amount respectively to 57.4, 60.3 and 62.8 kg.

Solution :

$$Q_1 = 57.4 \text{ kg}; \quad Q_3 = 62.8 \text{ kg}; \quad Q_2 = 60.3 \text{ kg}.$$

$$\therefore Sk = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)} = \frac{(62.8 - 60.3) - (60.3 - 57.4)}{62.8 - 57.4} = -0.074.$$

Example 6.5.2.

Calculate Pearson's 2nd coefficient of skewness for a frequency distribution of Differential Aptitude Test scores, having a mean of 98.69, an SD of 14.08 and a median of 98.25.

Solution :

$$\bar{X} = 98.69; \quad s = 14.08; \quad Mdn = 98.25.$$

$$\therefore Sk = \frac{3(\bar{X} - Mdn)}{s} = \frac{3(98.69 - 98.25)}{14.08} = +0.094.$$

6.6 KURTOSIS

Kurtosis denotes the degree of peakedness of a frequency distribution, compared to that of the normal distribution. The normal distribution is said to be *mesokurtic* and its peakedness of a medium order is taken as the standard (Fig. 6.4). A *leptokurtic* distribution has a higher and sharper peak, thicker tails and a narrower body than the normal distribution. Thus, a leptokurtic

distribution has higher frequencies of scores near its centre and at its two tails than a normal distribution with the same mean and variance, but has lower frequencies of scores of intermediate magnitudes. Student's *t* distributions are generally leptokurtic unless the sample is very large. A *platykurtic* distribution is flatter at its centre, broader in the body and thinner at the tails than the normal distribution because compared to the latter, the former carries lower frequencies of scores near its centre and at its tails, but higher frequencies of scores of intermediate magnitudes.

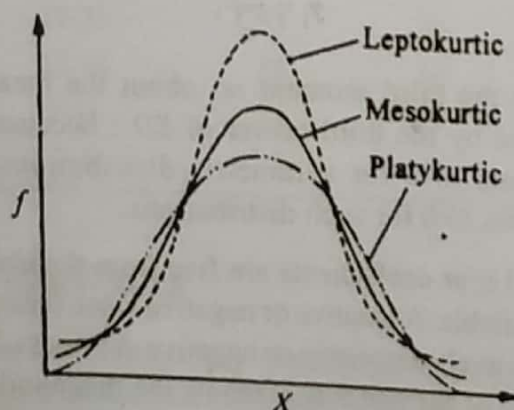


Fig. 6.4. Different forms of kurtosis.

Measures of kurtosis

(a) *Percentile coefficient of kurtosis (κ) :*

$$\kappa = \frac{P_{75} - P_{25}}{2(P_{90} - P_{10})}; \quad \text{or, } \kappa = \frac{Q}{P_{90} - P_{10}}.$$

For mesokurtic, platykurtic and leptokurtic distributions, κ amounts respectively to 0.263, > 0.263 , and < 0.263 ; the higher the value above 0.263, the greater is the platykurtosis while the lower it is below 0.263, the greater is the

leptokurtosis. Because the normal distribution is mesokurtic, κ amounts to 0.263 for it. On the contrary, t distributions are leptokurtic and have κ values lower than 0.263.

(b) *Moment coefficient of kurtosis (γ_2) :*

$$\gamma_2 = \frac{m_4}{m_2^2} - 3; \quad \text{or, } \gamma_2 = \frac{m_4}{s^4} - 3;$$

where s is the *SD*, and m_2 and m_4 are respectively the second and fourth central moments. For mesokurtic, platykurtic and leptokurtic distributions, γ_2 amounts to zero, a negative value and a positive value, respectively. Thus, γ_2 amounts to zero for the mesokurtic normal distribution, but has positive values for leptokurtic t distributions.

Example 6.6.1.

Compute the percentile coefficient of kurtosis for a frequency distribution of Differential Aptitude Test scores, having a 10th percentile of 80.60, a 25th percentile of 88.89, a 75th percentile of 108.15 and a 90th percentile of 116.03.

Solution :

$$P_{10} = 80.60; \quad P_{25} = 88.89; \quad P_{75} = 108.15; \quad P_{90} = 116.03.$$

$$\therefore \kappa = \frac{P_{75} - P_{25}}{2(P_{90} - P_{10})} = \frac{108.15 - 88.89}{2(116.03 - 80.60)} = 0.272.$$

As κ exceeds 0.263, the distribution is slightly platykurtic.

6.7 STUDENT'S t DISTRIBUTION

The probability distribution of the scores (X) of a small sample ($n < 30$), drawn from a normally distributed population, differs from the normal distribution, but conforms to a probability distribution called the *Student's t distribution*. The latter was originally derived in 1907 by W.S. Gossett writing under the pseudonym of "Student". The t distribution is the probability distribution of the statistic t which, like z , is a standard score obtained by the linear transformation of any raw score (X) by dividing its difference from the mean with the *SD*.

$$t = \frac{X - \mu}{\sigma}.$$

A sample statistic such as the sample mean may also be linearly transformed into a t score by dividing the difference between the statistic and the corresponding population parameter, with the *SE* of that statistic.

$$t = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}.$$

Like the scores of a small sample, the statistics of small samples, drawn from a normally distributed population, have probability distributions conforming to t distributions.

The difference between two sample statistics can also be transformed into a t score by dividing the deviation of that difference from the difference between the respecting parameters, with the *SE* of the difference between those statistics. Where \bar{X}_1 and \bar{X}_2 are the means of two samples drawn from two normally distributed populations having μ_1 and μ_2 as the respective parametric means, $\sigma_{\bar{X}_1 - \bar{X}_2}$ is the *SE* of the difference between \bar{X}_1 and \bar{X}_2 , \hat{s} is the pooled or common *SD* of the two samples, and n_1 and n_2 are the respective sample sizes,

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}},$$

$$\text{or, } t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\hat{s} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

If the samples have been drawn from the same population,

$$\mu_1 = \mu_2; \quad \therefore t = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}_1 - \bar{X}_2}}.$$

The probability distribution of such differences between two sample statistics also conforms to the t distribution.

Properties

1. t distributions are *continuous probability distributions for small samples*. The statistic t is thus a *small-sample statistic*.

2. t distributions can be worked out *theoretically* using the *Gossett's equation*. Where Y_o is a constant depending on the degrees of freedom (df) of the given t score, Y is the probability of random occurrence of that t score, π is the constant ratio between circumference and diameter of a circle, the Gossett equation expresses Y as a function of t in the following manner.

$$Y_o = \frac{[(df-1)/2]!}{[(df-2)/2]!} \times \frac{1}{\sqrt{\pi df}};$$

$$Y = \frac{Y_o}{\left[1 + \frac{t^2}{df}\right]^{(df+1)/2}}.$$

The theoretical t distribution can be graphically drawn as a t distribution curve by plotting the computed Y scores, scaled along the ordinate axis, against the respective t scores scaled along the abscissa axis (Fig. 6.5).

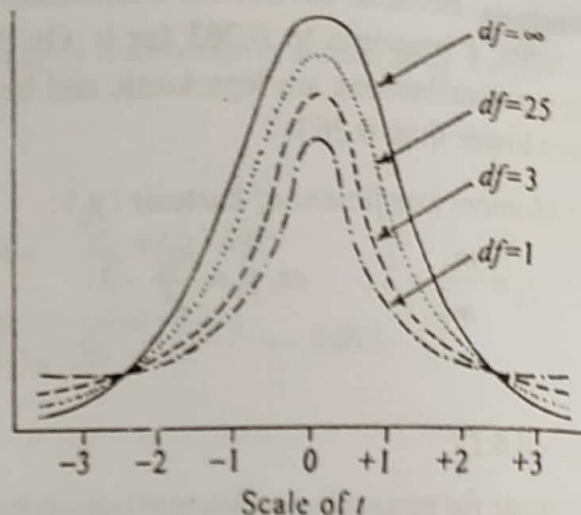


Fig. 6.5. Some t distributions for different degrees of freedom.

3. The t distribution forms a *bilaterally symmetrical* curve around a single central peak and possesses *no skewness*. The mean, median and mode coincide with the centre. Thus, the t distribution has a mean of 0 which is the t score for μ . The standard deviation of the distribution amounts to $\sqrt{n/(n-2)}$, where $(n-2)$ is a positive integer.

4. There are different t distributions for the t scores having different *degrees of freedom*. With the rise in df (and hence, in the sample size), the t distribution progressively approaches the normal distribution in shape — there are negligible differences between the two distributions when n equals or exceeds 30, and almost no difference when n approaches ∞ . For large samples, the equation for t distribution resembles that for the unit normal curve :

$$Y = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

Hence, though a small-sample statistic, t is applicable even to large samples because of the close identity of normal and t distributions for large samples.

5. In contrast to the normal distribution, t distributions are *leptokurtic*, having percentile coefficients of kurtosis lower than 0.263.

Leptokurtosis rises with the fall in the df and hence, with the reduction in the sample size. When df approaches infinity, the decline in leptokurtosis causes the t distribution to become mesokurtic and to coincide with the normal distribution.

6. Like the normal distribution, t distributions are asymptotic with their tails extending to $-\infty$ and $+\infty$.

7. Because probability distributions are distributions of relative frequencies, the total area under any t distribution curve, representing the total frequency n , is taken as 1.00.

8. The fractional area of a t distribution curve between the ordinates at any two t scores on the X axis gives that proportion of the total cases which falls between those given t scores. However, as t distributions vary in shape according to the df , such a range between the t scores as includes a specific fractional area like 0.95, varies with the df .

9. When a t score has been computed for the difference between two sample means, the sum of the fractional areas of the relevant t distribution in its two tails, beyond respectively the negative and positive values of the computed t , gives the probability P of obtaining by mere chance the observed difference between the sample means, irrespective of the algebraic sign of the difference.

10. A critical t score (t_α) is the highest t score with a given df , upto which the observed results (e.g., an observed difference between two sample means) have a specified probability P of occurring by mere chance. It is also the t score beyond which lies that specific fractional area in the tail(s) of the t distribution for a given df , as equals the chosen significance level (α) (§ 7.3 and 7.5, Table 7.1).

The t distributions have wide applications in finding the significance of the observed results and in computing the confidence intervals of population means (See Table 6.7).

6.8 CONFIDENCE INTERVALS

The confidence interval of a parameter is the range of scores within which the parameter has a given probability of lying. This probability is called the *fiducial probability*. The two scores (L_1 and L_2), forming respectively the lower and upper limits of the confidence interval, are called its lower and upper *confidence (fiducial) limits*. The fiducial probability is called the *confidence level* when expressed as a percentage, and gives the degree of confidence in expecting the parameter to lie in the given confidence interval. The confidence interval is an *interval estimate* of the parameter and is more dependable than a *point estimate* (page 11).

Confidence interval for mean

Confidence limits are computed for a parametric mean (μ) using the critical z or t score and the standard error ($\sigma_{\bar{X}}$) of the mean.

(a) Using z scores :

Means (\bar{X}) of large samples ($n \geq 30$) have an almost normal sampling distribution around μ irrespective of whether or not the variable has a normal distribution in the population, provided the latter has a finite variance (page 83). So, 95% of the sample means lies in the interval $\mu \pm 1.96\sigma_{\bar{X}}$ where 1.96 is the z score at the tail end of the

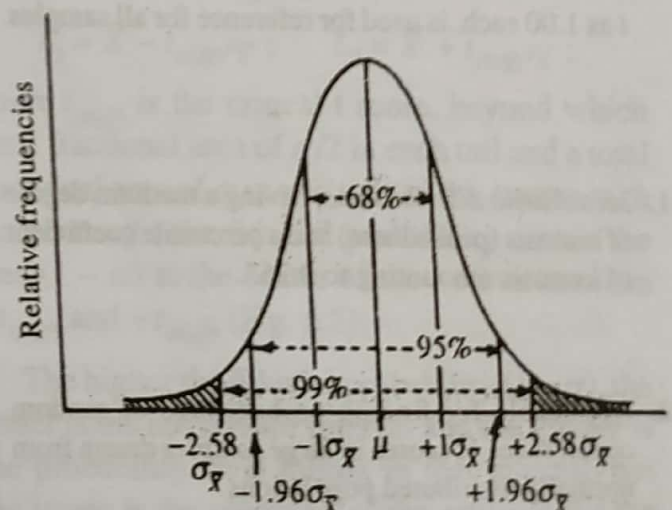


Fig. 6.6. Normal curve showing fractional areas for some confidence intervals.

Table 6.7. Comparison of normal and *t* distributions.

Normal distributions	Student's <i>t</i> distributions
1. <i>Continuous</i> probability distributions for scores of continuous variables.	1. <i>Continuous</i> probability distributions for scores of continuous variables.
2. Can be worked out <i>theoretically</i> using the <i>Gaussian equation</i> .	2. Can be worked out <i>theoretically</i> using the <i>Gossette equation</i> .
3. <i>Unimodal</i> distributions.	3. <i>Unimodal</i> distributions.
4. <i>Asymptotic tails</i> , progressively nearing the abscissa or X axis, but not reaching the latter before $-\infty$ or $+\infty$.	4. <i>Asymptotic tails</i> , progressively nearing the abscissa or X axis, but not reaching the latter before $-\infty$ or $+\infty$.
5. <i>Mean</i> of the distribution amounts to 0.00 which is the <i>z</i> score for μ .	5. <i>Mean</i> of the distribution amounts to 0.00 which is the <i>t</i> score for μ .
6. The <i>ordinate at the mean</i> ($z = 0.00$) is the highest of all the ordinates, indicating the probability of the highest number of occurrences of this score — thus, the mean coincides with the mode of the distribution.	6. The <i>ordinate at the mean</i> ($t = 0.00$) is the highest of all the ordinates so that the mean coincides with the mode of the distribution.
7. Perfect bilateral symmetry — <i>no skewness</i> ; skewness coefficients have zero values.	7. Perfect bilateral symmetry — <i>no skewness</i> ; skewness coefficients have zero values.
8. The ordinate at the mean ($z = 0.00$) bisects the distribution into two halves of equal areas and identical shapes — the mean thus coincides with the median.	8. The area of the distribution is divided into two equal halves of identical shape by the ordinate at the mean ($t = 0.00$) — the mean thus equals the median.
9. Distributions vary with sample size (n), SD (σ) and class interval size (i).	9. Distributions vary with the <i>degrees of freedom</i> of <i>t</i> scores, approaching the normal distribution with the rise in <i>df</i> .
10. A <i>unit normal curve</i> , computed by taking n , σ and i as 1.00 each, is used for reference for all samples.	10. Any computed <i>t</i> score has to be referred to the particular <i>t</i> distribution for the specific <i>df</i> of the computed <i>t</i> . For large samples, however, the equation for <i>t</i> distribution resembles that for the unit normal curve.
11. <i>Mesokurtic</i> distributions, having a medium degree of kurtosis (peakedness) and a percentile coefficient of kurtosis amounting to 0.263.	11. <i>Leptokurtic</i> distributions with sharper peaks, narrower bodies and thicker tails than mesokurtic distributions, and having percentile coefficients of kurtosis below 0.263 — leptokurtosis declines with the rise in <i>df</i> .
12. Applicable for finding probabilities of random occurrences of scores in <i>large samples</i> drawn from normally distributed populations.	12. Applicable for finding probabilities of random occurrences of scores in <i>small samples</i> drawn from normally distributed populations; applicable in case of <i>large samples</i> also, because of close identity of normal distribution with <i>t</i> distributions for large samples.

fractional area of 0.4750 in each half of the unit normal curve (Fig. 6.6). The probability is, therefore, 0.95 that the difference $(\bar{X} - \mu)$ ranges between $-1.96\sigma_{\bar{X}}$ and $+1.96\sigma_{\bar{X}}$. So, using the unbiased $s_{\bar{X}}$ of the sample, the fiducial probability amounts to 0.95 that μ lies in the interval $\bar{X} \pm 1.96s_{\bar{X}}$.

Similarly, 99% of the sample means lies in the interval $\mu \pm 2.58\sigma_{\bar{X}}$ where 2.58 is the z score at the tail end of the fractional area of 0.4951 in each half of the unit normal curve (Fig. 6.6). So, the fiducial probability amounts to 0.99 that μ lies in the interval $\bar{X} \pm 2.58s_{\bar{X}}$.

Thus, for the fiducial probability of $(1 - \alpha)$, the two confidence limits are worked out as follows using a random sample drawn *with replacement* :

$$L_1 = \bar{X} - z_{\alpha} s_{\bar{X}} ; \quad L_2 = \bar{X} + z_{\alpha} s_{\bar{X}} ;$$

where z_{α} is the critical z score, beyond which lies a fractional area of $\alpha/2$ in each tail and a total fractional area of α in two tails of the unit normal curve — the fractional area from μ to z_{α} evidently amounts to $(1 - \alpha)/2$ in each half of the normal curve. Thus, for a fiducial probability $(1 - \alpha)$ of 0.95,

$$L_1 = \bar{X} - 1.96s_{\bar{X}} ; \quad L_2 = \bar{X} + 1.96s_{\bar{X}} ;$$

similarly, for a fiducial probability of 0.99,

$$L_1 = \bar{X} - 2.58s_{\bar{X}} ; \quad L_2 = \bar{X} + 2.58s_{\bar{X}} .$$

Using a sample drawn *without replacement* from a finite population of size N ,

$$L_1 = \bar{X} - z_{\alpha} s_{\bar{X}} \sqrt{\frac{N-n}{N-1}} ;$$

$$L_2 = \bar{X} + z_{\alpha} s_{\bar{X}} \sqrt{\frac{N-n}{N-1}} .$$

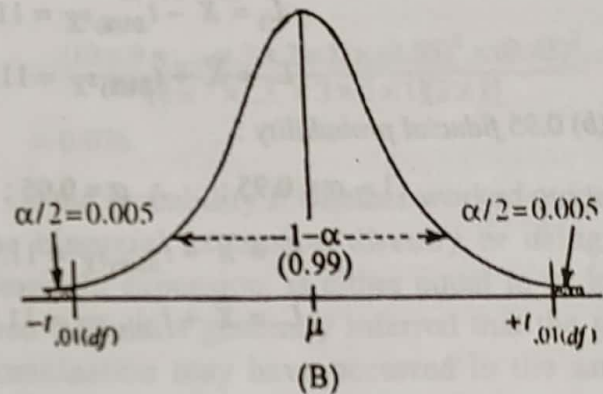
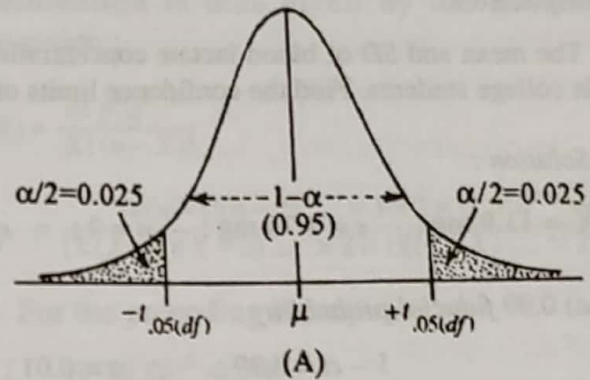


Fig. 6.7. t distribution showing fractional areas $(1 - \alpha)$ for (A) 95% and (B) 99% confidence intervals.

(b) Using t scores :

Irrespective of the sample size, the confidence limits for a fiducial probability of $(1 - \alpha)$ can be computed using the critical t score with the given df .

$$L_1 = \bar{X} - t_{\alpha(df)} s_{\bar{X}} ; \quad L_2 = \bar{X} + t_{\alpha(df)} s_{\bar{X}} ;$$

where $t_{\alpha(df)}$ is the critical t score, beyond which lies a fractional area of $\alpha/2$ in each tail and a total fractional area of α in two tails of the t curve with the given df ; the fiducial probability equals the area $(1 - \alpha)$ at the centre of the t curve between $-t_{\alpha(df)}$ and $+t_{\alpha(df)}$ (Fig. 6.7).

The higher the fiducial probability $(1 - \alpha)$, the wider is the confidence interval and the greater is the probability of μ falling in that interval, but the lower is the precision of the estimate for the true value of μ .

Example 6.8.1.

The mean and *SD* of blood lactate concentration amounted respectively to 11.9 and 1.76 mg dL⁻¹ for nine male college students. Find the confidence limits of the population mean at 0.95 and 0.99 fiducial probabilities.

Solution :

$$\bar{X} = 11.9 \text{ mg}; \quad s = 1.76 \text{ mg}; \quad n = 9; \quad s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{1.76}{\sqrt{9}} = 0.587 \text{ mg}; \quad df = n - 1 = 9 - 1 = 8.$$

(a) 0.99 fiducial probability :

$$1 - \alpha = 0.99; \quad \therefore \alpha = 0.01; \quad t_{\alpha(df)} = t_{.01(8)} = 3.355 \quad (\text{Table B}).$$

$$L_1 = \bar{X} - t_{.01(8)} s_{\bar{X}} = 11.9 - 3.355 \times 0.587 = 9.93 \text{ mg};$$

$$L_2 = \bar{X} + t_{.01(8)} s_{\bar{X}} = 11.9 + 3.355 \times 0.587 = 13.87 \text{ mg}.$$

(b) 0.95 fiducial probability :

$$1 - \alpha = 0.95; \quad \therefore \alpha = 0.05; \quad t_{\alpha(df)} = t_{.05(8)} = 2.306 \quad (\text{Table B}).$$

$$L_1 = \bar{X} - t_{.05(8)} s_{\bar{X}} = 11.9 - 2.306 \times 0.587 = 10.55 \text{ mg};$$

$$L_2 = \bar{X} + t_{.05(8)} s_{\bar{X}} = 11.9 + 2.306 \times 0.587 = 13.25 \text{ mg}.$$

Example 6.8.2.

The mean and *SD* of the scores obtained by 256 children in a digital-symbol learning test were found to be 82 and 32 respectively. Find the fiducial limits of the population mean at 99% confidence level.

Solution :

$$\bar{X} = 82; \quad s = 32; \quad n = 256; \quad s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{32}{\sqrt{256}} = 2.0.$$

$$1 - \alpha = 0.99; \quad \therefore (1 - \alpha)/2 = 0.4950.$$

The *z* score, having fractional area of 0.4950 between it and μ , amounts to 2.58 (Table A).

$$L_1 = \bar{X} - z_{\alpha} s_{\bar{X}} = 82 - 2.58 \times 2 = 76.84;$$

$$L_2 = \bar{X} + z_{\alpha} s_{\bar{X}} = 82 + 2.58 \times 2 = 87.16.$$

6.9 BINOMIAL DISTRIBUTION

If a population is divided into two classes with respect to a variable, the probabilities of occurrence of different combinations of events or individuals of the two classes in a sample, drawn from that population, are given by the series of terms of the binomial equation. Thus, the distribution of the probabilities of occurrence (the relative expected frequencies) of events or

individuals of either class can be computed theoretically from the binomial equation and constitute a *binomial probability distribution*.

Computation of binomial probabilities

(a) *Using the binomial expansion directly :*

Where *n* is the total number of events or cases in the sample, *p* and *q* are the proportions of the population in the two classes, and *q* equals (1 - *p*),

the binomial expansion may be written as follows:

$$\begin{aligned}(p + q)^n &= p^n + np^{n-1}q + \frac{n(n-1)}{1 \times 2} p^{n-2}q^2 \\ &+ \frac{n(n-1)(n-2)}{1 \times 2 \times 3} p^{n-3}q^3 + \dots \\ &+ \frac{n(n-1)(n-2) \dots \times 3 \times 2}{1 \times 2 \times 3 \times \dots \times (n-1)} pq^{n-1} + q^n.\end{aligned}$$

In this expansion, each term such as p^n , $np^{n-1}q$ and q^n gives the relative expected frequency or probability of occurrence of that specific combination of events or cases of two classes as is given by the respective powers of p and q in that term. The coefficients such as n , $n(n-1)/(1 \times 2)$, etc., of these terms are the *binomial coefficients* which can be directly obtained from the Pascal's triangle.

If, say, a sample of 10 individuals is drawn from a population with 0.55 and 0.45 proportions of cases in its two classes, the probability (P) of the random occurrence of a combination of 8 cases of the p class and 2 cases of the q class in the sample is given by the binomial expansion as follows:

$$n = 10; \quad p = 0.55; \quad q = 0.45;$$

$$\text{cases of } q \text{ class} = 2;$$

$$\text{cases of } p \text{ class} = n - 2 = 10 - 2 = 8;$$

$$\begin{aligned}P(8) &= \frac{n(n-1)}{1 \times 2} p^{n-2} q^2 \\ &= \frac{10 \times (10-1)}{1 \times 2} (0.55)^8 \times (0.45)^2 \\ &= 0.076.\end{aligned}$$

(b) Using the Bernoulli expansion:

The probability of a given combination of cases of the two classes (or of a given number of cases in one of the classes) in a binomial distribution is also given by the *Bernoulli distribution* formulated by Jacques Bernoulli (1654-1705), a Swiss mathematician. Where the given combination consists of X and $(n - X)$ numbers of cases of respectively p and q classes, and $n!$, $X!$ and $(n - X)!$ represent the factorials of n , X and $(n - X)$ respectively, the probability $P(X)$ of that

combination is thus given by the Bernoulli expansion:

$$\begin{aligned}P(X) &= \frac{n! p^X q^{n-X}}{X! (n-X)!} \\ &= \frac{[n(n-1)(n-2) \dots \times 3 \times 2 \times 1] p^X q^{n-X}}{[X(X-1)(X-2) \dots \times 2 \times 1][(n-X) \dots \times 2 \times 1]}\end{aligned}$$

For the preceding example,

$$\begin{aligned}P(8) &= \frac{10! (0.55)^8 \times (0.45)^2}{8! (10-8)!} \\ &= \frac{[10 \times 9 \times \dots \times 3 \times 2 \times 1] \times (0.55)^8 \times (0.45)^2}{[8 \times 7 \times \dots \times 3 \times 2 \times 1][2 \times 1]} \\ &= 0.076.\end{aligned}$$

If the probability P whether worked out using the binomial expansion directly or using the Bernoulli expansion, is either equal to or lower than 0.05, it is generally inferred that the given combination may have occurred in the sample owing to reasons other than random sampling ($P \leq 0.05$). But if P exceeds 0.05, it may be inferred that there is a high probability of the given combination occurring in the sample by mere chances of random sampling itself.

The binomial distribution of the probabilities of events of either class may be represented by a *bar diagram* (Fig. 6.8). Numbers of events of the specified class are scaled along the X axis and form the midpoints of the bases of the bars; heights of the bars are determined by the relative frequencies (probabilities) scaled along the Y axis. Because the events occur in whole numbers only, the bars are separated by intervening gaps to indicate the discrete nature of distribution.

A binomial distribution may also be represented by a *frequency polygon* (Fig. 6.9), plotting the relative frequencies (probabilities), scaled along the Y axis, against the respective number of events scaled along the X axis.

In problems like those involving sex-ratios, fertility rates, mortality rates, infection rates,

learning experiments and psychological judgements, the binomial distribution may give the probability of events.

The *absolute expected frequency* (f_e) of a given combination in a certain number of samples or trials is obtained by multiplying the relative expected frequency or probability, computed as above for a sample or trial, by the total number (k) of such samples or trials used. In the preceding example, the absolute expected frequency of the combination of 8 cases of p class and 2 cases of q class in 100 such samples of size 10 each, is given by :

$$f_e(X) = kP(X) = 100 \times 0.076 = 7.6.$$

Properties

1. The binomial distribution is a *discrete probability distribution*. It gives probabilities of whole numbers (0, 1, 2, 3, ..., n) of events or cases of a class ; because there cannot be any fractional occurrence of an event, the whole number of events form a discontinuous series with intervening gaps between them. This makes the distribution discrete.

2. It is the probability distribution of

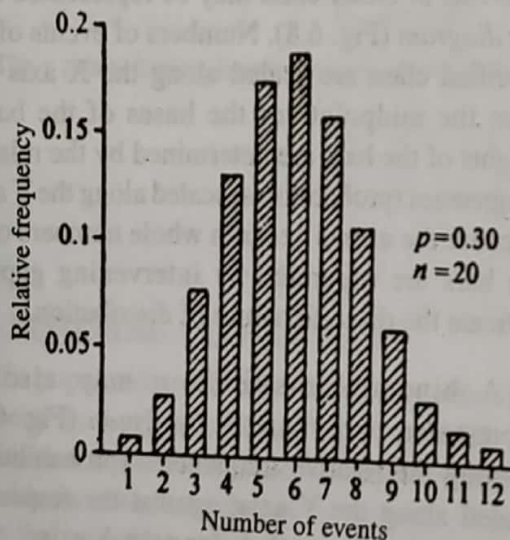


Fig. 6.8. Bar diagram of a binomial distribution of events belonging to the class with the proportion p in the population.

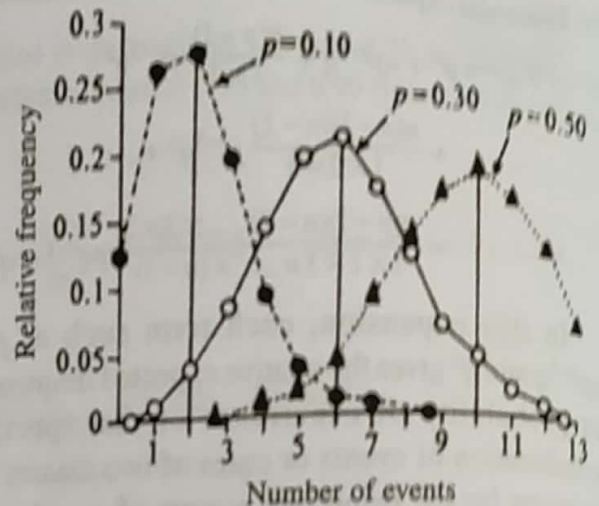


Fig. 6.9. Frequency polygons of binomial distributions of events of the class with the proportion p in the population, $n = 20$.

dichotomized variables divided into only two classes according to the presence or absence of some property. Such variables include sex, HIV positive-negative, success-failure, yes-no answers, and pregnant-nonpregnant.

3. Binomial probability distributions can be worked out *theoretically* on the basis of the *binomial equation* (see above). So, the probability of a given number of cases in either class, or of a given combination of cases in both classes of the variable, can be obtained from a specific term of the binomial expansion.

4. Neither of the two classes consists of rare cases so that neither of their respective proportions, p and q , in the population is too much lower than 0.50.

5. The events or cases of either class occur *at random* and are *independent* of each other.

6. *Mean* (μ), *variance* (σ^2), *SD* (σ) and the *coefficient of dispersion* (CD) of a binomial distribution of cases of the class having the proportion p in the population, are obtained from the sample size (n) and the proportions (p and q) of the cases in the two classes.

$$\mu = np ; \quad \sigma^2 = npq ; \quad \sigma = \sqrt{npq} .$$

$$CD = \frac{\sigma^2}{\mu} = \frac{npq}{np} = q.$$

7. *Skewness and kurtosis* of binomial distribution depend on the proportions (p and q) of the two classes in the population.

(a) The moment coefficient (γ_1) of skewness of a binomial distribution of the class having the proportion p in the population, is given by :

$$\gamma_1 = \frac{q-p}{\sqrt{npq}}.$$

Thus, the distribution is bilaterally symmetrical and has *no skewness* when $p = q = 0.50$. But it has a *negative skewness* if $p > 0.50$, and its left or low-value tail is more drawn out than the right one. On the contrary, it has a *positive skewness* with a longer right tail when $p < 0.50$.

(b) The binomial distribution is *platykurtic* so long as p lies between about 0.2114 and 0.7886, but turns *leptokurtic* if p is either below 0.2114 or above 0.7886. Its kurtosis is given as follows by the moment coefficient (γ_2) of kurtosis :

$$\gamma_2 = \frac{1-6pq}{npq}.$$

Example 6.9.1.

Work out and interpret the binomial probability of random occurrence of 8 male cockroaches in a sample of 10 drawn from a cockroach population with a male : female ratio of 45 : 55. Also find the absolute expected frequency of occurrence of a similar combination of males and females in 100 such samples of 10 cockroaches each.

Solution :

$$\text{proportion } (p) \text{ of males} = \frac{45}{100} = 0.45; \text{ proportion } (q) \text{ of females} = \frac{55}{100} = 0.55;$$

$$\text{sample size } (n) = 10; \text{ number } (X) \text{ of males in sample} = 8; \text{ number } (k) \text{ of samples} = 100.$$

(a) *Using the binomial expansion :*

$$(p+q)^n = p^n + np^{n-1}q + \frac{n(n-1)}{1 \times 2} p^{n-2}q^2 + \dots + \frac{n(n-1)(n-2) \dots \times 3 \times 2}{1 \times 2 \times 3 \times \dots \times (n-1)} pq^{n-1} + q^n.$$

$$\begin{aligned} \text{or, } (0.45 + 0.55)^{10} &= (0.45)^{10} + 10 \times (0.45)^9 \times 0.55 + \frac{10 \times 9}{1 \times 2} \times (0.45)^8 \times (0.55)^2 + \dots \\ &\quad + \frac{10 \times 9 \times 8 \times \dots \times 3 \times 2}{1 \times 2 \times 3 \times \dots \times 8 \times 9} \times 0.45 \times (0.55)^9 + (0.55)^{10}. \end{aligned}$$

Assumptions

For applying the binomial distribution to the observed results, it should be justifiable to assume that :

(a) the population is *dichotomized* with an intervening gap between the two classes ;

(b) the events or cases of each class occur at *random and independent of each other* in the sample so that occurrence of one event does not influence the probability of occurrence of any other event ;

(c) the proportions of cases in the two classes of the population have remained *unchanged* during sampling and are known with reasonable accuracy ;

(d) the mean and variance of the distribution of events of the relevant class (p) are given by the following : $\mu = np$; $\sigma^2 = npq$.

(e) the *coefficient of dispersion* (CD) of the events of one class equals the proportion of cases in the other, and falls short of 1 : $CD = q < 1.00$.

The probability $P(8)$ of the combination of 8 males and 2 females in a sample is given by the third term of the expansion, which has 8 and 2 as the respective powers of p and q . Thus,

$$P(8) = \frac{10 \times 9}{1 \times 2} \times (0.45)^8 \times (0.55)^2 = 0.023.$$

(b) Using the Bernoulli expansion :

$$P(X) = \frac{n! p^X q^{n-X}}{X! (n-X)!},$$

$$\text{or, } P(8) = \frac{10! (0.45)^8 \times (0.55)^2}{8! (10-8)!} = \frac{(10 \times 9 \times \dots \times 2 \times 1) (0.45)^8 (0.55)^2}{(8 \times 7 \times \dots \times 2 \times 1) (2 \times 1)} = 0.023.$$

As the binomial probability P is found to be lower than 0.05, it may be inferred that the given number of 8 males has occurred in the sample of 10 owing to reasons other than random sampling ($P < 0.05$).

Absolute expected frequency (f_e) of the similar combination of males and females in k number of such samples ($k = 100$) is given by :

$$f_e(8) = kP(8) = 100 \times 0.023 = 2.3.$$

Example 6.9.2.

Find and interpret the probability of random occurrence of 7 goitre cases in a sample of 10 drawn from a population with 40% incidence of endemic goitre. What is the absolute expected frequency of occurrence of a similar combination of goitre and nongoitre cases in 250 such samples of 10 individuals each ?

Solution :

$$\text{proportion } (p) \text{ of goitre cases} = \frac{40}{100} = 0.40 ;$$

$$\text{proportion } (q) \text{ of nongoitre cases} = 1.00 - 0.40 = 0.60.$$

$$\text{sample size } (n) = 10 ; \text{ number of goitre cases } (X) \text{ in sample} = 7 ; \text{ number } (k) \text{ of sample} = 250.$$

(a) Using the binomial expansion :

$$(p + q)^n = p^n + np^{n-1}q + \frac{n(n-1)}{1 \times 2} p^{n-2}q^2 + \frac{n(n-1)(n-2)}{1 \times 2 \times 3} p^{n-3}q^3 + \dots$$

$$+ \frac{n(n-1)(n-2) \dots \times 3 \times 2}{1 \times 2 \times 3 \times \dots (n-2)(n-1)} p q^{n-1} + q^n,$$

$$\begin{aligned} \text{or, } (0.40 + 0.60)^n &= (0.40)^{10} + 10(0.40)^9 \times 0.60 + \frac{10 \times 9}{1 \times 2} \times (0.40)^8 \times (0.60)^2 \\ &+ \frac{10 \times 9 \times 8}{1 \times 2 \times 3} \times (0.40)^7 \times (0.60)^3 + \dots + \frac{10 \times 9 \times 8 \times \dots \times 3 \times 2}{1 \times 2 \times 3 \times \dots \times 8 \times 9} \times (0.40) \times (0.60)^9 \\ &+ (0.60)^{10}. \end{aligned}$$

The probability $P(7)$ of the combination of 7 goitre and 3 nongoitre cases is given by the fourth term of the expansion, which has 7 and 3 as the powers of p and q , respectively. Thus,

$$P(7) = \frac{10 \times 9 \times 8}{1 \times 2 \times 3} \times (0.40)^7 \times (0.60)^3 = 0.042.$$

(b) Using the Bernoulli expansion :

$$P(X) = \frac{n! p^X q^{n-X}}{X! (n-X)!},$$

$$\text{or, } P(7) = \frac{10! (0.40)^7 \times (0.60)^3}{7! (10-7)!} = \frac{(10 \times 9 \times \dots \times 2 \times 1) \times (0.40)^7 \times (0.60)^3}{(7 \times 6 \times \dots \times 2 \times 1) \times (3 \times 2 \times 1)} = 0.042.$$

As the binomial probability P is found to be lower than 0.05, it may be inferred that the given number of 7 goitre cases has occurred in the sample of 10 owing to reasons other than random sampling ($P < 0.05$).

Absolute expected frequency (f_e) of the similar combination of goitre and nongoitre cases in k number of such samples ($k = 250$) is given by :

$$f_e(7) = kP(7) = 250 \times 0.042 = 10.5.$$

6.10 POISSON DISTRIBUTION

In a population dichotomized with respect to a variable, one of the two classes may sometimes consist of rare events or cases forming a very low proportion of the population. The distribution of relative expected frequencies or probabilities of different numbers of such rare events in a sample from such a population may conform to a *theoretical probability distribution*, called the *Poisson distribution* after its formulator S. D. Poisson, a French mathematician. In such cases, the probabilities or relative expected frequencies of 0, 1, 2, 3, ..., n numbers of rare events, among the total number n of all the events, are given by the successive terms of the following Poisson distribution :

$$\frac{1}{e^\mu}; \frac{\mu}{1! e^\mu}; \frac{\mu^2}{2! e^\mu}; \frac{\mu^3}{3! e^\mu}; \dots \frac{\mu^n}{n! e^\mu};$$

where μ is the mean of the Poisson distribution, n is the sample size, and e is the base of the natural logarithm, approximating 2.7183. Using the sample mean \bar{X} , these probabilities may be stated in terms of the Poisson distribution in the

following series :

$$\frac{1}{e^{\bar{X}}}; \frac{\bar{X}}{1! e^{\bar{X}}}; \frac{\bar{X}^2}{2! e^{\bar{X}}}; \frac{\bar{X}^3}{3! e^{\bar{X}}}; \dots \frac{\bar{X}^n}{n! e^{\bar{X}}}.$$

Thus, the probability $P(X)$ of X number of rare events, occurring in a sample, is given by :

$$P(X) = \frac{\bar{X}^X}{X! e^{\bar{X}}}$$

$$= \frac{\bar{X}^X}{[X(X-1)(X-2) \dots \times 2 \times 1] (2.7183)^{\bar{X}}}$$

The absolute expected frequency (f_e) of a given number (X) of rare events in the total number (k) of samples, each of size n , is obtained as follows :

$$f_e(X) = kP(X).$$

Properties

1. The Poisson distribution is the *probability distribution of rare events* belonging to one of the two classes of a *dichotomized variable*. The class, whose events have a distribution in conformity to the Poisson distribution, has a very low or near-zero proportion (p) in the population ; the events of the other class occur far more frequently and have a proportion (q) close to 1 in the population.

2. It is a *discrete probability distribution*, because it is a probability distribution of whole numbers (0, 1, 2, 3, n) of events, separated by intervening gaps because of no possibility of occurrence of fractional numbers of rare events.

3. It can be formed theoretically by working out probabilities of random occurrences of the events of the rare class using the Poisson equation (see above).

4. The rare events occur at *random* and *independent of each other* in the sample – the probability of occurrence of one rare event is not increased or decreased by the occurrence of any other.

5. The rare events may occur either *spatially* in a specified space or volume (e.g., Down syndrome patients in a sample of children, or abnormal erythrocytes in a hemocytometer chamber) or *temporally* in a given time interval (e.g., number of mutations per day, or number of suicide cases in a month).

6. Mean (μ) and variance (σ^2) of a Poisson distribution of rare events are identical, finite and very low — less than 5 — compared to the total n number of events of both the classes ; thus, the

coefficient of dispersion (CD) is 1 or nearly 1 for a Poisson distribution.

$$\mu = np; \quad \sigma^2 = np; \quad CD = \frac{\sigma^2}{\mu} = 1.$$

7. Poisson distributions have positive skewness, given by the moment coefficient (γ_1) of skewness :

$$\gamma_1 = \frac{1}{\sqrt{\mu}} = \frac{1}{\sqrt{np}}.$$

For μ as low as 0.1, the distribution is reverse J-shaped with no left tail and a prolonged right tail. As μ rises, a peak appears, but the right or high-value tail is longer than the left one – this positive skewness declines progressively with the rise in μ .

8. Poisson distribution is *leptokurtic*, its leptokurtosis declining with the rise in μ . In terms of the moment coefficient (γ_2) of kurtosis,

$$\gamma_2 = \frac{1}{\mu} = \frac{1}{np}.$$

Because the shape of Poisson distribution depends on the mean, probabilities of rare events also depend on the latter.

Assumptions

For applying the Poisson distribution to the observed results, it should be justifiable to assume that :

(a) the variable under investigation is divided into only two classes with intervening gaps between the two ;

(b) the proportion p of the population, falling in the rare class whose events are under consideration, is very low or near-zero while the proportion q in the other class is correspondingly high ;

(c) μ and σ^2 of the probability distribution of rare cases are identical, finite, both equal to np and lower than 5, so that the CD is 1 or nearly 1 ;

(d) the rare events occur at *random* and *independent of each other* in the sample.

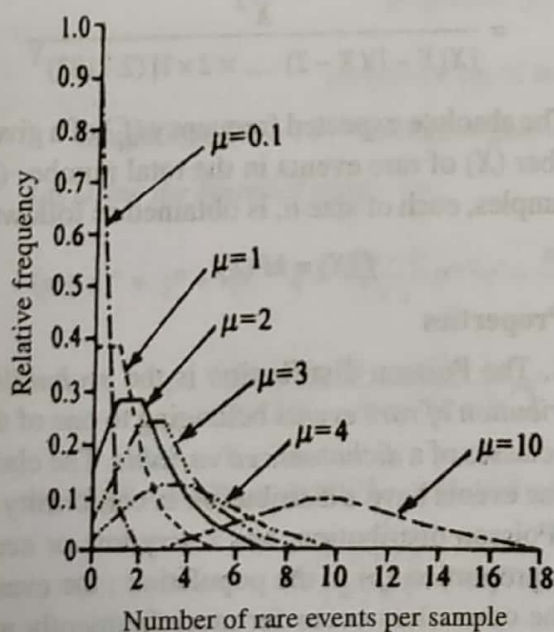


Fig. 6.10. Frequency polygons of Poisson distribution.

Table 6.8. Comparison between binomial and Poisson distributions.

Binomial distributions	Poisson distributions
1. Probability distributions of numbers of cases of either of the two classes of a <i>dichotomized variable</i> in samples.	1. Probability distributions of numbers of cases of one specific class of a <i>dichotomized variable</i> in samples.
2. Neither p nor q , the respective proportions of the two classes in the population, is much lower or higher than 0.50 so that neither can be considered a rare class.	2. Probability distributions of cases of only the <i>rare class</i> of dichotomized variables, which has a very low, even near-zero proportion p in the population.
3. <i>Discrete probability distributions</i> of only whole numbers of cases of either class, constituting a discontinuous series without any fractional number in between.	3. <i>Discrete probability distributions</i> of only whole numbers of cases of the rare class, constituting a discontinuous series with no fractional number.
4. Theoretical model of distribution can be worked out using the <i>binomial equation</i> .	4. Theoretical model of distribution can be worked out using the <i>Poisson equation</i> .
5. Distributions of cases of a class are bilaterally symmetrical with <i>no skewness</i> so long as its proportion p equals 0.50 in the population, but are <i>negatively skewed</i> when p exceeds 0.50, and <i>positively skewed</i> when p is less than 0.50 — the moment coefficient of skewness is given by : $(q - p)/\sqrt{npq}$.	5. Distribution of cases of the rare class are <i>positively skewed</i> , with the positive skewness declining with the rise in the mean (μ) — the moment coefficient of skewness is given by : $1/\sqrt{\mu}$.
6. Distributions of cases of the p class are <i>platykurtic</i> so long as p is between about 0.2114 and 0.7886, but <i>leptokurtic</i> when p is beyond that range — the moment coefficient of kurtosis is given by : $(1 - 6pq)/npq$.	6. Distributions of cases of the rare class are <i>leptokurtic</i> , the leptokurtosis declining with the rise in μ — the moment coefficient of kurtosis is given by : $1/\mu$.
7. The mean (μ) and the variance (σ^2) of the distribution of cases of the p class are given by np and npq respectively, where p and q are the respective proportions of the two classes in the population.	7. Both μ and σ^2 of the distribution of rare cases with a proportion p in the population are given by np and amount to less than 5.
8. The coefficient of dispersion (CD) of the distribution of the class with a proportion p in the population equals the proportion q of the other class and is consequently lower than 1.00.	8. Being the ratio of μ and σ^2 , the CD equals 1.00 for the distribution of the rare class.

Example 6.10.1.

Work out and interpret the probability of random occurrence of 5 Down syndrome cases in a sample of 200 children from a population having 0.5% incidence of that genetic disorder. Also compute the absolute expected frequency of such Down syndrome cases in 1000 such samples.

Solution :

Where p is the proportion of Down syndrome children in the population,

$$p = \frac{0.5}{100} = 0.005 ; \quad n = 200 ; \quad k = 1000 ; \quad \bar{X} = np = 200 \times 0.005 = 1.00 ;$$

$$\therefore P(X) = \frac{\bar{X}^X}{X! e^{\bar{X}}} ; \quad \text{or, } P(5) = \frac{1^5}{5! (2.7183)^1} = 0.003.$$

As the Poisson probability P is found to be lower than 0.05, it may be inferred that the given number of 5 Down syndrome cases has occurred in the sample of 200 owing to reasons other than mere chances of random sampling ($P < 0.05$).

Absolute expected frequency (f_e) of such cases in 1000 such samples ($k = 1000$) is given by :

$$f_e(X) = kP(X) = 1000 \times 0.003 = 3.$$

GLOSSARY

asymptotic tail : such a tail of a distribution as approaches ever nearer to the abscissa / ordinate but never meets it within a finite distance.

Bernoulli distribution : a probability distribution formulated by Jacques Bernoulli to help give the binomial probability of random occurrence of events or cases of either of the classes of a dichotomous variable in a sample.

binomial distribution : discrete probability distributions of cases or events belonging to either of the classes of a dichotomous variable, worked out theoretically from the binomial equation.

central limit theorem : means of large samples have an almost normal sampling distribution around the mean of their population even if the latter has a non-normal distribution of the relevant scores, provided the population has a finite variance.

central theorem of probability : a variable has an almost normal distribution if its scores depend on the independent effects of many other variables, acting at random and with no interaction with each other.

confidence interval of mean : that range of scores, computed from the sample mean, in which the population mean has a specified probability of falling.

Gaussian equation : the mathematical equation formulated by Karl Fredrich Gauss and used in working out a normal probability distribution theoretically.

kurtosis : the magnitude of peakedness of a probability (or frequency) distribution.

leptokurtosis : the peakedness of a distribution having a higher or sharper peak, a narrower body and thicker tails than the mesokurtic normal distribution.

mesokurtosis : the medium degree of peakedness of the normal distribution, generally used as a standard of peakedness of distributions.

normal distribution : continuous probability (or frequency) distributions which can be theoretically worked out using the Gaussian equation and are unimodal, bilaterally symmetrical and asymptotic.

normal distribution, best-fitting : a normal distribution that can fit best with an observed distribution, and has the same mean, SD and sample size as those of the latter.

- platykurtosis** : the peakedness of a distribution having a flatter peak, a broader body and thinner tails than the mesokurtic normal distribution.
- Poisson distribution** : discrete probability distributions of cases or events, belonging to the rare class of a dichotomous variable, worked out theoretically from the equation formulated by S.D. Poisson.
- probability** : the relative frequency of the events, cases or scores of a given type among a very large or almost infinite number of total events, cases or scores.
- probability distribution** : the distribution of probabilities of occurrences of different events, cases or scores of a given variable in a sample or a population.
- probability distribution, continuous** : probability distributions of scores of a continuous measurement variable.
- probability distribution, discrete** : probability distributions of events, cases or scores of a discontinuous variable.
- probability distribution, theoretical** : a probability distribution of the cases or scores of a variable, worked out theoretically on the basis of a specific mathematical equation or model.
- sampling theory of means** : the means of random samples from a normally distributed population form a normal sampling distribution around the population mean.
- skewness** : bilateral asymmetry of a probability (or frequency) distribution, with one tail of the latter longer than the other.
- skewness, negative** : bilateral asymmetry of a probability (or frequency) distribution with its negative or low-value tail longer than its positive or high-value tail.
- skewness, positive** : bilateral asymmetry of a probability (or frequency) distribution with its positive or high-value tail longer than its negative or low-value tail.
- t* distribution** : continuous probability distributions, which can be worked out theoretically using the equation formulated by W.S. Gossett ("Student") for the probabilities of scores in small samples drawn from a normally distributed population.
- t* score** : standard score in σ unit, worked out by linear transformation of a raw score of a small sample drawn from a normally distributed population.
- t* score, critical** : the *t* score beyond which lies that specific fractional area in the tail(s) of the *t* distribution for a given *df*, as equals the chosen level (α) of significance.
- z* score** : standard score in σ unit, worked out by linear transformation of a raw score of a variable distributed normally in the population.
- z* score, critical** : the *z* score beyond which lies that specific fractional area in the tail(s) of a unit normal curve as equals the chosen level (α) of significance.
- unit normal curve** : the normal probability distribution curve obtained by plotting the probability (*Y*) of each *z* score against the latter where the sample size (*n*), the class interval size (*i*) and the *SD* of raw scores are each taken as 1.00.

7. TESTING OF HYPOTHESIS

The significance (importance or meaningfulness) of the result of any experiment using a sample is assessed by estimating the probability (P) of such a result arising by chance owing to the drawing of the particular sample by random sampling. For estimating this probability of random occurrence of the observed result, a standard score — often z or t — is computed from the observed data and the probability of its random occurrence is then worked out using respectively the normal and t distributions.

7.1 DIFFERENCE BETWEEN MEANS

In an experiment to study the effect of an independent variable (page 5) on a particular dependent variable (page 4), two or more groups may be randomly sampled from a population and exposed to different levels (doses, amplitudes, intensities, etc.) of the independent variable. The dependent variable is then studied or measured in each group to find whether or not the groups differ from each other with respect to the mean of the dependent variable scores. There may, however, be two alternative reasons for any observed difference between the group means. *First*, the independent variable may not have produced any change in the dependent variable; even then, the mean dependent variable scores of two groups, exposed to different levels of the independent variable, may differ from each other due simply to different *sampling errors* (page 67) of the group means — two group means in such a case do not differ significantly, are different estimates of the same population mean for the dependent variable, and their difference is so small as can be accounted for by the *SE of difference between the means* (pages 72-73) because these groups still belong to the same population. *Alternatively*, the

independent variable may indeed have produced changes in the dependent variable; the difference between group means of the dependent variable, in such cases, is too large to be explained away by the *SE of difference* between the means or by the different sampling errors of the group means — the group means may then be considered to differ significantly and to be the estimates/representatives of the parametric means of two different populations, to which the two groups are considered as belonging.

In view of these two alternative possibilities, significance of the difference between two group means must be assessed in terms of the *SE of that difference*, viz., whether (i) the observed difference is *not significant*, being small enough to be explainable by the *SE of the difference* as being due simply to different sampling errors of those means, or (ii) the difference is *significant*, being too large to be explained away by the *SE of the difference* as arising merely from the sampling errors. For such assessment, the observed difference between two group means is generally transformed into a *standard score* (either z or t), using the *SE of that difference* (pages 74-75); the probability (P) of the computed standard score occurring by mere chance due to random sampling is then found out using either the unit normal curve or the t distribution, as the case may be (vide § 7.6 and 7.7). So long as this computed probability (P) does not exceed a particular chosen level of probability, viz., the level of significance (α , vide § 7.3), P is considered too low ($P \leq \alpha$) — it is then inferred that the probability P of the observed difference having arisen by mere chance of random sampling is too low and so, the observed difference between the means may be

considered *significant*. But if the computed P exceeds the chosen significance level ($P > \alpha$), P is considered too high — it is then inferred that the probability P of the random occurrence of the observed difference is too high and so, the group means *do not differ significantly*; in other words, the independent variable has not produced significant changes in the dependent variable.

7.2 NULL HYPOTHESIS

Experiments are generally performed with random samples instead of the entire population (pages 7-9); the inferences drawn from the results observed in samples are then sought to be generalized over the entire population. Thus, there is always a probability that the observed results may have arisen from the accidental choice of the particular sample drawn by random sampling in accordance with the laws of probability, and would not have been obtained if the entire population were subjected to the experiment instead of a random sample. So, in drawing an inference, the investigator has to weigh this probability (P) of the observed results having arisen from the chances associated with random sampling. In other words, in interpreting the result of any experiment *using samples*, the investigator has to assess the probability (P) of the correctness of a *null hypothesis* (H_0) which proposes to nullify or negate the hypothesis under investigation in the experiment, by professing that the observed result (i) has arisen by chance due to the drawing of the particular sample by random sampling, (ii) would not have been obtained if the entire population would have been used in the experiment instead of the sample, and (iii) has, therefore, *no significance*. On the contrary, the hypothesis investigated in the experiment and contested by the null hypothesis is called the *alternative hypothesis* (H_a). For drawing the inference, the result of an experiment has to be subjected to the

statistical testing of the contesting hypotheses, H_0 and H_a , relevant to that experiment.

The H_0 takes different forms according to the H_a that it contests. However, what is common to all forms of H_0 is the proposition that the observed results have been produced solely due to the accidental choice of the particular sample by the operation of probabilities inherent in random sampling. A few examples of the H_0 and H_a are given below.

(a) In testing the significance of a *difference between the means* of two groups (samples), the H_0 proposes that the observed difference is *not significant*, that it has resulted from mere chances of random sampling, and that the difference would have been zero if the entire population were used in the experiment instead of groups sampled at random. In other words, the H_0 contends that both the groups (samples) belong to the same population, their means serve as estimates of the same parametric mean, and the observed difference between the means is due simply to their different sampling errors. In contrast, the H_a being tested in the experiment proposes that the two groups (samples) belong to two different populations, that their means are estimates of two different parametric means (μ_1 and μ_2), and that there is a *significant difference* between the group means, which cannot be explained away by the *SE* of such differences. Thus,

$$H_a : \mu_1 \neq \mu_2 ; \quad H_0 : \mu_1 = \mu_2$$

(b) For the significance of a *correlation* between two variables, the H_0 proposes that there is *no significant correlation* between the variables, that the observed correlation has resulted from the accidental sampling of a particular group depending on laws of probability, and that the correlation coefficient would have amounted to zero if the entire population, instead of a random sample, were studied in the experiment. On the contrary, the

H_a proposes that there is a *significant correlation* between the variables — or, in other words, there is a significant difference between the observed correlation coefficient and zero.

(c) In testing the *goodness of fit* between a frequency distribution observed in the experiment and a proposed distribution based on normal, binomial or Mendelian 9 : 3 : 3 : 1 distribution, the H_0 contends that the observed distribution does not differ significantly from the proposed distribution, that any difference found between the two has resulted from the use of a particular sample or group drawn by chance by random sampling, and that the two distributions would have a *significant goodness of fit* if the entire population were studied instead of a random sample. In contrast, the H_a proposes that there is a significant difference between the two distributions, which cannot be explained away by mere chances associated with random sampling.

It should be borne in mind that even in case of the same experiment, there would be different null hypotheses for one-tail and two-tail tests, respectively (vide § 7.5).

Although an experiment is performed to

inquire into a specific alternative hypothesis (H_a), the latter is not directly tested statistically; on the contrary, its acceptance or rejection is determined by the rejection or retention of the corresponding null hypothesis (H_0). For this, the H_0 is tested statistically by working out the probability P of getting the observed difference, correlation or association by mere chance owing to random sampling. Stated otherwise, the probability P of the H_0 being correct is worked out statistically. If this estimated P does not exceed a particular chosen level of probability (*level of significance* or α), the probability of correctness of the H_0 is considered too low ($P \leq \alpha$); the H_0 may then be rejected, the H_a may instead be accepted, and the observed difference, correlation or association is considered *significant* (vide § 7.3). But if the estimated P exceeds the chosen α , the probability of correctness or propriety of the H_0 is considered too high ($P > \alpha$); in that case, the H_0 is retained, the H_a cannot be accepted, and the observed results are *not significant*.

7.3 LEVELS OF SIGNIFICANCE

A *level of significance* (α) is that probability of the random (chance) occurrence

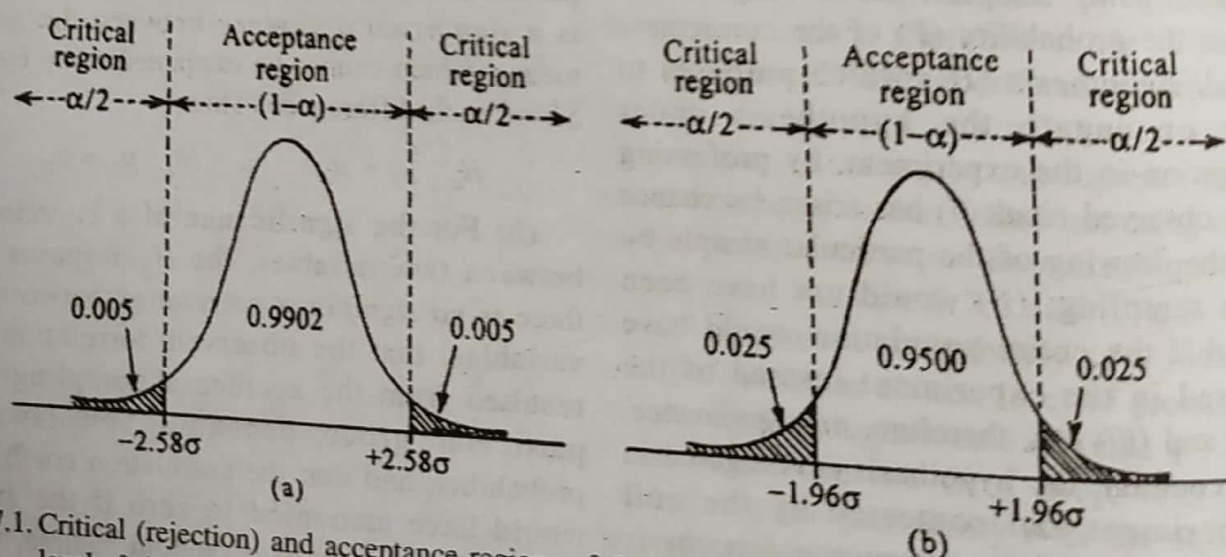


Fig. 7.1. Critical (rejection) and acceptance regions of the normal distribution for (a) 0.01 level and (b) 0.05 level of significance. Critical $z_{0.01} = 2.58\sigma$, and $z_{0.05} = 1.96\sigma$.

of observed results, upto and below which the probability P of correctness of the null hypothesis is considered *too low*. Thus, so long as the estimated P does not exceed the chosen α ($P \leq \alpha$), the H_0 is *rejected* because of the low probability (P) of its correctness, the H_a is accepted, and the results of the experiment are considered *significant*. But if the estimated P exceeds α ($P > \alpha$), the probability of the H_0 being correct is considered *too high* so that the H_0 cannot be rejected; so the H_0 is retained, the H_a cannot be accepted, and the observed results are consequently adjudged as *not significant*.

$P \leq \alpha$: H_0 rejected ; H_a accepted ;
results significant.

$P > \alpha$: H_0 retained ; H_a rejected ;
results not significant.

For example, the investigator may use an α of 0.05 for the interpretation of the observed results of an experiment. In such a case, the H_0 is rejected and the observed results are considered significant if the probability P of getting the results by mere chance, due to random sampling, works out to be 0.05 or less ($P \leq 0.05$) — this means that the results would be adjudged *significant* if out of 100 such trials, only 5 or less number of times the observed results may arise merely from the accidental choice of the sample (group) by random sampling. But if P is found here to exceed the chosen α of 0.05 ($P > 0.05$), the H_0 would be retained because of the too-high probability of its correctness, and the observed results would be adjudged *not significant*. For biological experiments, α is generally fixed either at the 0.05 level or at a still lower level such as 0.02, 0.01 and 0.001.

The level of significance (α) is given by the fractional area(s) in the tail(s) of the normal or t distribution beyond the relevant critical z or t score (pages 82-83 and 91). It may be recalled (pages 82-83) that the sum of the fractional

areas beyond a two-tail critical z_α or t_α in *both tails* of the relevant probability distribution equals the corresponding *two-tail significance level* (2-tail α) and constitutes the *two-tail critical (rejection) region* of the distribution (Fig. 7.1 and 7.4a); thus, the fractional area beyond the two-tail critical z_α or t_α in each tail of the distribution corresponds to $\alpha/2$. So long as the computed z or t score in a *two-tail test* (vide § 7.5) is not lower than the two-tail critical z_α or t_α , the fractional area beyond that computed z or t score does not exceed the critical region; because the total critical region in both tails corresponds to α and the total fractional area in both tails beyond the computed z or t corresponds to the probability (P) of the correctness of H_0 , P does not exceed the chosen α ($P \leq \alpha$) as long as the computed z or t either equals or exceeds the critical z_α or t_α , and the H_0 may be *rejected*. The remaining area in the central part of the normal or t distribution, extending between the critical scores in the two tails of the latter, constitutes the *acceptance region* ($1 - \alpha$) of the distribution (Fig. 7.1). If the computed z or t score is lower than the critical z_α or t_α , that computed z or t falls in the acceptance region; the sum of the fractional areas beyond the computed z or t in both tails, corresponding to P , now exceeds the sum (α) of the critical regions in the two-tails and encroaches into the acceptance region so that $P > \alpha$, warranting the *acceptance* of the H_0 .

It may also be recalled (page 82) that the fractional area beyond the critical z_α or t_α in a *single tail* of the normal or t distribution equals the corresponding *one-tail significance level* (1-tail α) and constitutes the entire *one-tail rejection region* of the distribution (Fig. 7.4b). In a *one-tail test* (vide § 7.5), the H_0 is rejected if the computed z or t score either equals or exceeds the one-tail critical z_α or t_α and consequently, falls in the one-tail rejection region ($P \leq \alpha$). On the contrary, the H_0 is

accepted in a one-tail test if the computed z or t is lower than the critical z_α or t_α and consequently, falls in the acceptance region $(1 - \alpha)$ of the distribution; here, the acceptance region $(1 - \alpha)$ of the distribution ranges from the critical score in one of the tails and over the total remaining area of the distribution including its entire other tail (Fig. 7.4b).

7.4 ERRORS OF INFERENCE

Because the inference of an experiment is drawn through either the rejection or the acceptance of the H_0 according to the probability (P) of its correctness, the inference always suffers from probabilities of errors of inference owing to either (i) a wrongful rejection or (ii) a wrongful acceptance of the H_0 depending on the estimated P and the chosen α .

Type I error of inference

This consists of the *wrongful rejection of a true H_0* . In other words, it is the error made in accepting the accidental and insignificant results of an experiment as significant. As the H_0 is rejected if the probability (P) of its correctness does not exceed the significance level (α), the probability of the type I error is limited to the α used in making the inference. Stated otherwise, α is that maximum probability of type I error, which the investigator risks in rejecting the H_0 for making

the inference. Thus, the higher the chosen α for considering the H_0 untenable, the greater is the probability of the type I error. If the level of α used in making the inference amounts to 0.01, the type I error has a probability of occurring once in 100 such trials due to random sampling; if the α used is 0.05, the probability of type I error increases to five out of 100 trials.

In making any inference, the status of the probability (P) of the correctness of the H_0 should always be mentioned in relation to the significance level (α) used; e.g., $P < 0.01$, $P = 0.02$, $P < 0.05$, $P > 0.05$, $P < 0.001$, etc. This immediately gives out the probability of the type I error in the inference made.

Type II error of inference

This consists of a *wrongful acceptance of a false H_0* . It is thus the error committed in rejecting the really significant results of an experiment as insignificant. Stated otherwise, the type II error (β) is the failure in identifying the genuine results of an experiment and consequently rejecting a true H_a . Important factors in the type II error are as follows.

(a) Type II error (β) depends on the overlap between the distribution proposed by the H_a around its parametric mean (μ_a) and the H_0 distribution around the parametric mean (μ_0) of the latter (Fig. 7.2). Indeed, β is that fractional area of the H_a distribution which falls within

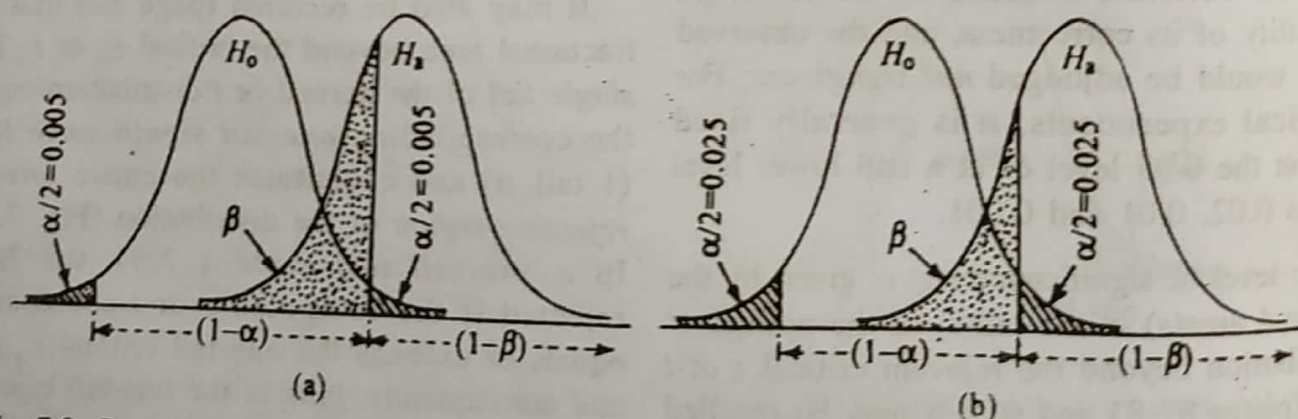


Fig. 7.2. Inverse relations between type I and type II errors. (a) Two-tail α : 0.01. (b) Two-tail α : 0.05.

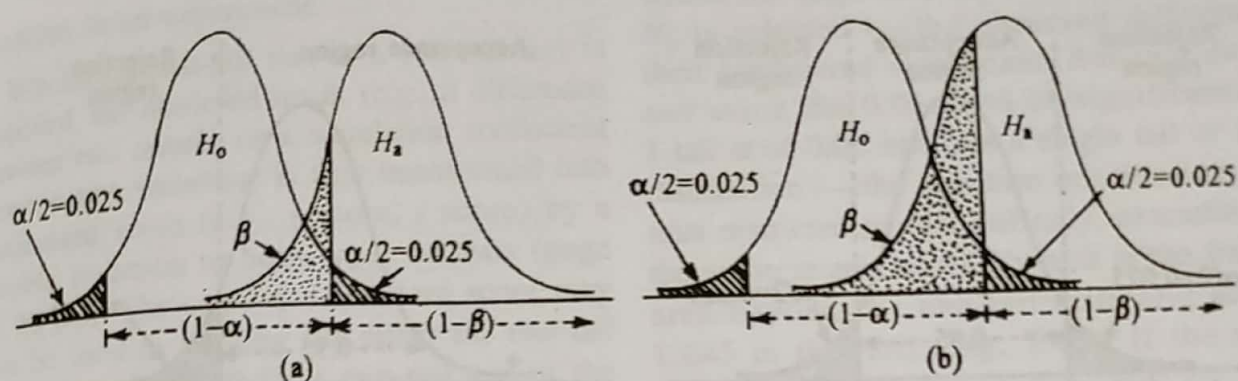


Fig. 7.3. Higher type II error (β) when the parametric means of H_0 and H_a distributions are closer (as in b) than when they are wider apart (as in a).

the acceptance region $(1 - \alpha)$ of the H_0 distribution.

(b) The lower the significance level (α) used in rejecting the H_0 , the narrower is the rejection region of the H_0 distribution and the lower is the probability of the type I error; but simultaneously, the wider is the acceptance region $(1 - \alpha)$, thus increasing the overlap between the H_a distribution and the latter, and enhancing the probability of the type II error or β correspondingly (Fig. 7.2a). Reverse is the case when α is increased. Thus, there is an inverse relationship between the probabilities of type I and type II errors (Fig. 7.2). Generally, a smaller α like 0.01, 0.02 and 0.001 may be chosen in spite of a higher risk of β , when it is intended to limit the risk of a wrong positive inference due to the acceptance of the H_a .

(c) The closer the parametric means (μ_0 and μ_a) of the H_0 and H_a distributions, the greater is the overlap of the two distributions (Fig. 7.3). This makes it more difficult to discriminate μ_0 from μ_a , and consequently enhances the probability of the type II error (β) of a wrongful acceptance of a false H_0 . On the contrary, β declines if μ_a and μ_0 are wider apart.

7.5 ONE-TAIL AND TWO-TAIL TESTS

Either a two-tail or a one-tail statistical test has to be used according to the H_0 in question.

Two-tail test

A two-tail test is a *nondirectional* statistical test for finding the significance of only the magnitude of the observed difference between the means of two groups (samples), irrespective of the algebraic sign of that difference. For example, if the mean serum cholesterol amounts to 220 mg dL⁻¹ in a group of diabetics and 180 mg dL⁻¹ in a group of nondiabetics, a two-tail t test may find whether or not there is a significant difference between the group means, without considering whether one of the group means is significantly higher (or lower) than the other. The H_0 proposes here that there is *no significant difference* between the group means, any observed difference having resulted from chances associated with random sampling. Since both positive and negative differences between the group means are under consideration here, the differences may lie in both tails of the distribution. Thus, the rejection region of the H_0 distribution involves both tails, amounting to $\alpha/2$ in each (Fig. 7.4a). The sum of the fractional areas in both tails of the H_0 distribution thus gives the two-tail probability (P) of correctness of the H_0 . The inference of such a two-tail test is limited to the existence/absence of any significant difference between the group means and does not take into consideration whether or not one of the means is higher or lower than the other.

variables in an experiment.

Whether a two-tail test or a one-tail test is proposed, the observed result (e.g., a difference between two means, or a correlation coefficient between two variables) is first transformed into a standard score (e.g., z score, t score) by a process common to both types of tests (page 75, 89-90). The computed standard score may next be used in working out either the two-tail probability (2-tail P) for a two-tail test or the one-tail probability (1-tail P) for a one-tail test (page 82). In a two-tail test, the observed result (e.g., the observed difference between two means) is considered significant, only if the computed P does not exceed the chosen two-tail significance level ($P \leq \alpha$). Similarly in a one-tail test, the observed result (e.g., a negative $\bar{X}_1 - \bar{X}_2$ difference) is considered significant, only if the computed P does not exceed the chosen one-tail significance level or 1-tail α ($P \leq \alpha$).

For a *two-tail test* using the 0.05 level of significance, i.e., the 2-tail α of 0.05, each tail of the H_0 distribution ends with a rejection or critical region ($\alpha/2$) of area 0.025, extending beyond the two-tail critical z score of 1.96 in that tail (Fig. 7.4a). If the z score computed from an observed ($\bar{X}_1 - \bar{X}_2$) difference lies either at or beyond -1.96 or $+1.96$, it falls

within the respective rejection regions and the H_0 is rejected — the observed difference is then considered significant. But in a *one-tail test* using the 0.05 level of significance, the 1-tail α of 0.05 involves a single tail of the H_0 distribution — the rejection or critical region is thus restricted to that tail only, extending over the entire α area of 0.05 which is the fractional area beyond the one-tail critical z score of 1.645 in that tail (Fig. 7.4b). If the z score computed from the observed ($\bar{X}_1 - \bar{X}_2$) difference amounts to 1.645 or more, it falls in the single-tail rejection region and the H_0 may consequently be rejected. It is thus evident that with an identical α , an observed difference may be significant in a one-tail test, but may fail to be significant in a two-tail test — a computed z score of 1.645 in the above example is significant for a one-tail test at the 0.05 level of significance, but is not significant for a two-tail test at the same 0.05 level.

Some one-tail and two-tail critical scores for different significance levels are tabulated in Table 7.1.

7.6 SIGNIFICANCE OF DIFFERENCE BETWEEN MEANS USING Z SCORES

For an experiment to study the effects of an *independent variable* on a particular *dependent*

Table 7.1. Some critical z , t and χ^2 scores.

Statistic	df	$\alpha = 0.05$		$\alpha = 0.01$		$\alpha = 0.005$	$\alpha = 0.001$
		1-tail	2-tail	1-tail	2-tail	1-tail	2-tail
z	—	1.645	1.960	2.327	2.575	2.575	3.290
t	∞	1.645	1.960	2.326	2.576	2.576	3.291
t	60	1.671	2.000	2.390	2.660	2.660	3.460
t	30	1.697	2.042	2.457	2.750	2.750	3.646
t	23	1.714	2.069	2.500	2.807	2.807	3.767
t	9	1.833	2.262	2.821	3.250	3.250	4.781
χ^2	3	6.251	7.815	9.837	11.341	11.341	16.268
χ^2	2	4.605	5.991	7.824	9.210	9.210	13.815
χ^2	1	2.706	3.841	5.412	6.635	6.635	10.827

variable in a given population (pages 4-5), often separate groups of individuals are drawn from that population by random sampling, and each group is treated with or exposed to a specific level, i.e., a particular amount, concentration, amplitude, intensity or qualitative type, of the independent variable. Frequently, one of the groups, called the *control group*, is given a level of treatment that is free from the independent variable while the others, called the *experimental groups*, receive different levels of the treatment containing specific doses or amounts of the independent variable. Subsequently, the dependent variable is estimated or studied in all the groups. Such an experiment, using groups consisting of separate sets of individuals or cases, is called an *independent group experiment*.

For drawing inference about the significance of the observed difference between the means of two large independent groups, each not less than 30 in size ($n \geq 30$), the *unit normal curve* (Table A of Appendix) is used in working out the probability (P) of the correctness of the H_0 , because the means of large groups, drawn from a normally distributed population, as well as their differences have sampling distributions conforming to the normal probability distribution (page 83). For this, the observed difference between the group means, say $(\bar{X}_1 - \bar{X}_2)$, is transformed into the standard z score and the probability (P) of occurrence of the computed z score by mere chance of random sampling is worked out using the unit normal curve areas. The H_0 is rejected and the observed difference between the means, $\bar{X}_1 - \bar{X}_2$, is considered significant, only if the P thus worked out does not exceed the chosen level of significance ($P \leq \alpha$).

For using any statistical test, the variable under investigation, the relevant population and the obtained data must fulfil certain criteria. It is, however, not always necessary to work out

whether the required criteria have actually been fulfilled; it will suffice if it is justifiable to assume that the required criteria have been fulfilled. Such criteria are thus called the *assumptions* for the test.

Assumptions for tests using z scores

For finding the significance of the observed result using z scores, it should be justifiable to assume that :

(a) the dependent variable is a *continuous* measurement variable ;

(b) its scores have a *normal distribution* in the population ;

(c) each score occurs in a group (sample) at *random and independent of all other scores* — this ensures the representative nature of the groups for the population so that the results obtained with the groups can be generalized over the population ;

(d) the groups (samples) are *large enough* ($n \geq 30$) so that their means and consequently, the differences between the means have normal sampling distributions ;

(e) the groups initially possess *homogeneous variances*, i.e., their variances are initially different estimates of the same population variance (*homoscedasticity*), differing only due to the sampling errors.

It follows from these assumptions that tests using the unit normal curve and z scores cannot be used if (i) the dependent variable is a discontinuous, ordinal or nominal variable, (ii) its scores have a non-normal or skewed distribution in the population, and (iii) if the groups are of small sizes ($n < 30$).

Two-tail test

The *null hypothesis* for a two-tail test proposes that there is *no significant difference* between the group means, that the group means are estimates of the same population mean, that

the observed difference between the means has resulted from mere chances of random sampling and would not be there if the entire population were used instead of the randomly sampled groups. Thus, if μ_1 and μ_2 are the population means estimated respectively by \bar{X}_1 and \bar{X}_2 of the two groups,

$$H_0 : \mu_1 - \mu_2 = 0 ; \quad H_a : \mu_1 - \mu_2 \neq 0.$$

The probability P of the correctness of this H_0 is worked out and interpreted as follows.

(a) The observed difference ($\bar{X}_1 - \bar{X}_2$) between the two group means is converted to z score, starting with the assumptions that the H_0 is correct, and μ_1 and μ_2 are identical. Where s_1 and s_2 are the SDs of the respective groups, $s_{\bar{X}_1}$ and $s_{\bar{X}_2}$ are the SEs of the respective group means, $s_{\bar{X}_1 - \bar{X}_2}$ is the SE of the difference between means, ($\mu_1 - \mu_2$) is zero, and n_1 and n_2 are the respective group sizes,

$$s_{\bar{X}_1} = \frac{s_1}{\sqrt{n_1}} ; \quad s_{\bar{X}_2} = \frac{s_2}{\sqrt{n_2}} ;$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{(s_{\bar{X}_1})^2 + (s_{\bar{X}_2})^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} ;$$

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} ;$$

$$\text{or, } z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

(b) the fractional area of the unit normal curve, extending from its mean (μ) to the computed z score, is taken from the *Area* column of the unit normal curve table (Table A of Appendix) and used in working out the two-tail probability (2-tail P) of the H_0 being correct (page 82).

$$P = 2 [0.5000 - (\text{fractional area from } \mu \text{ to the computed } z)].$$

(c) The P thus worked out is next compared with the chosen level of significance (α). If P is lower than the chosen α , the probability of chance occurrence of the observed difference, as proposed by the H_0 , is considered too low ($P < \alpha$); if P equals the chosen α , then also the probability of chance occurrence of ($\bar{X}_1 - \bar{X}_2$) is considered too low ($P = \alpha$); in both these cases, the H_0 is rejected, the H_a is accepted and the observed difference ($\bar{X}_1 - \bar{X}_2$) is considered *significant*. But if P exceeds the chosen α , the probability of the H_0 being correct is considered too high ($P > \alpha$); so, H_0 is retained in this case and the observed difference ($\bar{X}_1 - \bar{X}_2$) is considered *not significant*.

One-tail test

The null hypothesis for a one-tail test proposes that one group mean is not significantly higher (alternatively, lower) than another, that the observed result, showing one mean higher (or lower) than the other, have arisen by chance due to random sampling, and that no such result would have arisen if the entire population were studied instead of the randomly sampled groups. Thus,

$$\text{either, } H_0 : \mu_1 \geq \mu_2 ; \quad H_a : \mu_1 > \mu_2 ;$$

$$\text{or, } H_0 : \mu_1 \leq \mu_2 ; \quad H_a : \mu_1 < \mu_2.$$

The probability P of the correctness of the H_0 is worked out and interpreted as follows.

(a) The observed difference ($\bar{X}_1 - \bar{X}_2$) between the two group means is converted to z score in the same way as in the case of the two-tail test, starting with the assumption that the H_0 is correct.

$$z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(s_{\bar{X}_1})^2 + (s_{\bar{X}_2})^2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

(b) From the *Area* column of the unit normal curve (Table A of Appendix, the

fractional area of the unit normal curve from its μ to the computed z is taken and used in working out the one-tail probability (1-tail P) of correctness of the H_0 (pages 81-82).

$P = 0.5000 -$ (fractional area from μ to the computed z).

(c) If the P thus worked out is found to be either lower than or equal to the chosen

significance level ($P \leq \alpha$), it is considered too low, the H_0 is rejected and one group mean is considered *significantly higher (alternatively, lower)* than the other. But if P exceeds the chosen α ($P > \alpha$), it is considered too high, the H_0 is retained, and one group mean is *not significantly higher (alternatively, lower)* than the other.

Example 7.6.1.

The mean time for regeneration of amputated basal discs amounted to 18.6 hrs (SD 2.20 hrs) in 84 wild (nonmutated) animals of *Hydra vulgaris*, and 17.7 hrs (SD 2.70 hrs) in 86 somatic mutants of that species. (a) Find whether or not there is a significant difference between the mean regeneration times of the two groups ($\alpha = 0.05$). (b) Is the mean regeneration time significantly higher in the wild hydra than in the mutants ($\alpha = 0.01$)?

Solution :

Wild animals : $\bar{X}_1 = 18.6$ hrs ; $s_1 = 2.20$ hrs ; $n_1 = 84$.

Mutants : $\bar{X}_2 = 17.7$ hrs ; $s_2 = 2.70$ hrs ; $n_2 = 86$.

1. Two-tail test for the significance of the difference between means :

H_0 proposes that the two means, \bar{X}_1 and \bar{X}_2 , do not differ significantly, the observed difference being due only to chances of random sampling.

(a) Assuming the H_0 to be correct, the difference ($\bar{X}_1 - \bar{X}_2$) between the means is first transformed into the z score.

$$s_{\bar{X}_1} = \frac{s_1}{\sqrt{n_1}} = \frac{2.20}{\sqrt{84}} = 0.24 ; \quad s_{\bar{X}_2} = \frac{s_2}{\sqrt{n_2}} = \frac{2.70}{\sqrt{86}} = 0.29 ;$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{(s_{\bar{X}_1})^2 + (s_{\bar{X}_2})^2} = \sqrt{0.24^2 + 0.29^2} = 0.376 \text{ hr.}$$

$$z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{18.6 - 17.7}{0.376} = 2.39.$$

$$\text{Alternatively, } z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{18.6 - 17.7}{\sqrt{\frac{(2.20)^2}{84} + \frac{(2.70)^2}{86}}} = 2.39.$$

(b) The two-tail probability P of the computed z score occurring by mere chance of random sampling, is then worked out using the unit normal curve table (Table A of Appendix).

$$P = 2 [0.5000 - \text{(fractional area of unit normal curve from its } \mu \text{ to the computed } z \text{ of 2.39)}]$$

$$= 2 [0.5000 - 0.4916] = 0.017.$$

$$\alpha = 0.05$$

So, the probability P of 0.017 for obtaining the observed difference $(\bar{X}_1 - \bar{X}_2)$ or its z score by mere chance of random sampling is less than the chosen α of 0.05 and is thus considered too low. Hence, the H_0 is rejected and the difference between the means is considered significant ($P < 0.05$).

2. One-tail test for the significance of \bar{X}_1 being higher than \bar{X}_2 :

H_0 contends that \bar{X}_1 is not significantly higher than \bar{X}_2 . The difference $(\bar{X}_1 - \bar{X}_2)$ is converted to the z score in the same way as for the two-tail test, given above. Thus, $z = 2.39$.

P is worked out for a one-tail test, using the unit normal curve table (Table A of Appendix).

$$P = 0.5000 - (\text{area of the unit normal curve from its } \mu \text{ to the computed } z \text{ of } 2.39) \\ = 0.5000 - 0.4916 = 0.0084.$$

$$\alpha = 0.01.$$

The computed P of 0.0084 is considered too low as it is lower than the chosen α of 0.01. The H_0 is, therefore, rejected and the mean regeneration time is considered significantly higher in the wild *Hydra* than in the mutants ($P < 0.01$).

Example 7.6.2.

In a modified form of Differential Aptitude Test, the mean score obtained by a sample of 374 girls amounted to 98.7 (SD 14.08) while the mean score of another sample of 255 boys was 95.5 (SD 13.02). Is there any significant difference between the mean scores of the two sexes ($\alpha = 0.01$) ?

Solution :

A two-tail test is undertaken to find the probability P of correctness of the H_0 which contends that there is no significant difference between the two sample (group) means, the observed difference being due only to chances associated with random sampling.

$$\text{For girls : } \bar{X}_1 = 98.7; \quad s_1 = 14.08; \quad n_1 = 374.$$

$$\text{For boys : } \bar{X}_2 = 95.5; \quad s_2 = 13.02; \quad n_2 = 255.$$

(a) The difference between the two means is converted to the z score.

$$s_{\bar{X}_1} = \frac{s_1}{\sqrt{n_1}} = \frac{14.08}{\sqrt{374}} = 0.728; \quad s_{\bar{X}_2} = \frac{s_2}{\sqrt{n_2}} = \frac{13.02}{\sqrt{255}} = 0.815.$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{(s_{\bar{X}_1})^2 + (s_{\bar{X}_2})^2} = \sqrt{(0.728)^2 + (0.815)^2} = 1.093.$$

$$\left[\text{or, } s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(14.08)^2}{374} + \frac{(13.02)^2}{255}} = 1.093. \right]$$

$$z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{98.7 - 95.5}{1.093} = 2.93.$$

(b) The unit normal curve table (Table A of Appendix) is used to find the two-tail probability P of obtaining the computed z score, and hence the observed difference between the means, by chance due to random sampling.

$$P = 2 [0.5000 - (\text{area of the unit normal curve from its } \mu \text{ to the computed } z \text{ of } 2.93)] \\ = 2 [0.5000 - 0.4983] = 0.0034.$$

(c) Because the computed P of 0.0034 is lower than the chosen α of 0.01, the probability of the H_0 being correct is considered too low. The H_0 is, therefore, rejected and the mean DAT scores are considered significantly different in the two sexes ($P < 0.01$).

Example 7.6.3.

The mean systolic blood pressure amounted to 120.7 mm Hg (SD 25.90) in 40 normal young men and 139.5 mm Hg (SD 30.05) in 40 young men suffering from renal ischemia. (i) Is there a significant difference in the mean systolic blood pressure between normal individuals and ischemia patients? (ii) Is the mean significantly higher in renal ischemia patients? ($\alpha = 0.01$).

Solution :

For ischemia patients : $\bar{X}_1 = 139.5$ mm Hg; $s_1 = 30.05$ mm Hg; $n_1 = 40$.

For normal men : $\bar{X}_2 = 120.7$ mm Hg; $s_2 = 25.90$ mm Hg; $n_2 = 40$.

1. Two-tail test for significance of difference between means :

The H_0 contends that there is no significant difference between \bar{X}_1 and \bar{X}_2 , and the observed difference has resulted from the chances associated with random sampling.

(a) The difference $(\bar{X}_1 - \bar{X}_2)$ between the means is transformed into z score.

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(30.05)^2}{40} + \frac{(25.90)^2}{40}} = 6.273 \text{ mm Hg.}$$

$$z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{139.5 - 120.7}{6.273} = 3.00.$$

(b) The unit normal curve table (Table A of Appendix) is used to find the two-tail probability P of obtaining the computed z score by chance due to random sampling.

$$P = 2 [0.5000 - (\text{area of unit normal curve from its } \mu \text{ to the computed } z \text{ of } 3.00)] \\ = 2 [0.5000 - 0.4987] = 0.0026.$$

$\alpha = 0.01$.

(c) Because the computed P of 0.0026 is lower than the chosen α of 0.01, the probability of the H_0 being correct is considered too low. The H_0 is, therefore, rejected and the mean systolic BP is considered to differ significantly between normal and ischemic men ($P < 0.01$).

2. One tail-test for the significance of \bar{X}_1 being higher than \bar{X}_2 :

The H_0 proposes in this case that \bar{X}_1 is not significantly higher than \bar{X}_2 .

(a) The difference $(\bar{X}_1 - \bar{X}_2)$ is converted to z score in the same way as for the two-tail test, given above. Thus, $z = 3.00$.

(b) The one-tail probability P of obtaining the computed z by chance is worked out, using the unit normal curve table (Table A).

$$P = 0.5000 - (\text{area of the unit normal curve from its } \mu \text{ to the computed } z \text{ of } 3.00) \\ = 0.5000 - 0.4987 = 0.0013.$$

(c) As the computed P of 0.0013 is lower than the chosen α of 0.01, the probability of the H_0 being correct is too low. The H_0 is, therefore, rejected and the mean systolic BP of ischemic patients is considered to be significantly higher than that of normal individuals ($P < 0.01$).

7.7 t TESTS FOR SIGNIFICANCE OF DIFFERENCE BETWEEN MEANS

In experiments using small groups or samples ($n < 30$) drawn at random from a population in which the dependent variable under investigation has a normal distribution, the scores of small groups are distributed in the form of a t distribution. So, to test the significance of the difference between the means of two small groups, their difference $(\bar{X}_1 - \bar{X}_2)$ is converted to Student's t score which is then interpreted using the t distribution.

Assumptions for t tests

To apply the t test, it should be justifiable to assume that :

(a) the dependent variable whose changes are being studied, is a *continuous* measurement variable ;

(b) the variable has a *normal distribution* in the population ;

(c) each score of the dependent variable occurs *at random and independent of all other scores* in the group (sample) ;

(d) the groups have been sampled from population(s) having homogeneous variances (*homoscedasticity*) so that initially the variances of the groups differ merely due to their sampling errors.

It follows that t tests cannot be used (i) if the dependent variable is a discrete, ordinal or nominal variable, (ii) if it has a non-normal or skewed distribution in the population, or (iii) if the populations, from which the groups have been initially sampled, do not obey homoscedasticity.

Degrees of freedom

Because t distributions vary with the degrees of freedom (df) of the t scores, the computed t must be referred for interpretation to the t distribution specific for the df of the computed t .

Because the unit normal curve closely approximates or almost coincides with the t distributions for high degrees of freedom (i.e., large sample sizes), t test can be applied not only to small samples ($n < 30$) but also to large ones, even to those with n approaching ∞ . This is evident from the very close similarities between the critical t scores ($df = \infty$) and the critical z scores for identical levels of significance (Table 7.1). Thus, for finding the significance of difference between the means of two large groups or samples ($df = \infty$), either the z score may be computed and then referred to the unit normal curve table, or the t score may be computed and then compared with the critical t scores ($df = \infty$) for different levels of significance. But t tests alone are applicable to small groups because their

distributions then conform to *leptokurtic* t distributions only, instead of the mesokurtic normal distribution.

Two-tail t test :

This is undertaken to find the significance of an observed difference between two means, irrespective of the algebraic sign of the difference. It investigates the probability P of correctness of the null hypothesis (H_0) which proposes that the two means are not significantly different from each other and are estimates of the same population parameter (§ 7.5).

$$H_0 : \mu_1 = \mu_2 ; \quad H_a : \mu_1 \neq \mu_2.$$

Assuming the H_0 to be correct, the observed difference ($\bar{X}_1 - \bar{X}_2$) between the sample means is transformed into t score using the same basic formula as is used in computing z score (§ 7.6).

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}.$$

However, $s_{\bar{X}_1 - \bar{X}_2}$ may be computed here in several alternative ways according to the nature of the group or sample and the type of experiment.

The probability P of the H_0 being correct is obtained from the fractional areas in the two tails of the relevant t distribution beyond the computed t score. For this, the computed t score is compared with the *two-tail critical t score* having the same df as that of the computed t , and for the chosen level of significance (α). So long as the computed t exceeds or equals the critical t score, it falls within the rejection region (α) of the H_0 distribution beyond the critical t_α score so that the P given by the total fractional area in the two tails beyond the computed t does not exceed α ($P \leq \alpha$). The H_0 is then considered to have too low a probability of being correct and may be rejected — the means are then

considered to *differ significantly* from each other. But if the computed t is lower than the critical t score, the P given by the total fractional area in the two tails beyond the computed t exceeds the chosen α ($P > \alpha$); the H_0 is then taken to have a high probability of being correct and cannot be rejected — the difference between the means is then considered *not significant*.

One-tail t test :

This is used to explore whether or not one of the means is significantly higher (alternatively, lower) than the other. It explores the P of correctness of the H_0 which contends that one mean is not significantly higher (alternatively, lower) than the other.

$$\begin{aligned} \text{Either, } H_0 : \mu_1 \geq \mu_2 ; \quad H_a : \mu_1 > \mu_2 ; \\ \text{or, } H_0 : \mu_1 \leq \mu_2 ; \quad H_a : \mu_1 < \mu_2 . \end{aligned}$$

The observed difference ($\bar{X}_1 - \bar{X}_2$) between the means is converted to t score in the same way as in the two-tail test. The computed t score is compared with the *one-tail critical t score* for the chosen α and having the same df . If the computed t either exceeds or equals the critical t , the fractional area in one tail beyond the computed t , corresponding to the 1-tail P , does not exceed the rejection region (α) beyond the critical 1-tail t_α in that tail ($P \leq \alpha$). The H_0 is then considered to have a too low P and may be rejected — one of the means is then considered *significantly higher (or, lower)* than the other mean. But if the computed t is lower than the critical t , the fractional area (P) in one tail beyond the computed t exceeds the rejection region (α) beyond the critical t_α ($P > \alpha$); the H_0 cannot be rejected as the P is considered too high — one of the means is then *not significantly higher (or, lower)* than the other.

1. t tests for independent groups

Two or more groups of individuals, used in an *independent group experiment*, are randomly sampled from the population independent of

each other so that they consist of separate sets of individuals and may or may not be identical in size. After the groups have been exposed to or treated with different levels (doses, concentrations, qualities, etc.) of the independent variable, the dependent variable being investigated is measured in all the groups. The scores of the dependent variable, thus obtained from different independent groups, constitute *unpaired observations* and are *uncorrelated* to each other. The significance of the difference between the means of two such groups is found by *t* tests in different ways according as the groups are large or small, and equal or unequal in size.

(a) *For small and unequal-size independent groups :*

For two independent groups of unequal sizes ($n_1 \neq n_2$), either or both being small in size (< 30), the observed difference ($\bar{X}_1 - \bar{X}_2$) between their means is converted to a *t* score using a *pooled SD* (\hat{s}) or a *common variance* (\hat{s}^2). Where s_1^2 and s_2^2 are the variances of the respective groups, $s_{\bar{X}_1}$ and $s_{\bar{X}_2}$ are the SEs of the respective means, and $s_{\bar{X}_1 - \bar{X}_2}$ is the SE of the difference between the means,

$$\hat{s} = \sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{(s_{\bar{X}_1})^2 + (s_{\bar{X}_2})^2} = \sqrt{\frac{\hat{s}^2}{n_1} + \frac{\hat{s}^2}{n_2}}$$

$$= \hat{s} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} ; \quad \text{or, } t = \frac{\bar{X}_1 - \bar{X}_2}{\hat{s} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}} ;$$

$$\text{or, } t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2} \times \frac{n_1 + n_2}{n_1 n_2}}}$$

$$df = n_1 + n_2 - 2.$$

(b) *For both small and large independent groups of equal size :*

If two independent groups possess the same size (n), \hat{s} and *t* may be computed as follows, irrespective of the group sizes.

$$\hat{s} = \sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{2(n-1)}} = \sqrt{\frac{s_1^2 + s_2^2}{2}}$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{(s_{\bar{X}_1})^2 + (s_{\bar{X}_2})^2} = \sqrt{\frac{\hat{s}^2}{n} + \frac{\hat{s}^2}{n}} = \hat{s} \sqrt{\frac{2}{n}}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} ; \quad \text{or, } t = \frac{\bar{X}_1 - \bar{X}_2}{\hat{s} \sqrt{\frac{2}{n}}} ;$$

$$\text{or, } t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n(n-1)}}}$$

$$\text{or, } t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2 + s_2^2}{n}}}$$

$$df = 2(n - 1).$$

(c) *For large groups of unequal sizes :*

Where both the groups have large but unequal sizes ($n_1 > 30, n_2 > 30 ; n_1 \neq n_2$), the *t* score is computed for ($\bar{X}_1 - \bar{X}_2$) using the SDs of the individual groups, instead of the pooled SD.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}} ; \quad df = n_1 + n_2 - 2.$$

Interpretation of computed t : (i) The *t* score, computed by any above method is compared with critical *t* scores, having the same *df* as that of the computed *t*, but different levels of significance. (ii) Either two-tail or one-tail critical *t* scores are used in this

comparison, according as a two-tail or a one-tail t test is intended. (iii) The algebraic sign of the computed t is ignored in this comparison. (iv) If the computed t either equals or exceeds the critical t for a chosen level of significance (α) not exceeding 0.05, the H_0 is rejected and

the observed results are considered significant, respectively at or below that level ($P \leq \alpha$). But (v) if the critical t for a chosen α exceeds the computed t , the H_0 cannot be rejected and the observed results are not considered significant ($P > \alpha$).

Example 7.7.1.

The heights (cm) of 25 male and 20 female college students are presented in the first two columns of Table 7.2. Find if there is a significant difference between the mean heights of male and female college students.

Solution :

The H_0 proposes that there is no significant difference between the means. To find the probability P of the H_0 being correct, a two-tail t test is done using the pooled SD (\bar{s}) because the groups are small and of unequal sizes.

Table 7.2. Table for computing means and standard deviations of body height data.

Heights (cm)		$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	$X_2 - \bar{X}_2$	$(X_2 - \bar{X}_2)^2$
Males (X_1)	Females (X_2)				
163	164	- 3.4	11.56	+ 2.9	8.41
165	155	- 1.4	1.96	- 6.1	37.21
170	160	+ 3.6	12.96	- 1.1	1.21
162	154	- 4.4	19.36	- 7.1	50.41
160	160	- 6.4	40.96	- 1.1	1.21
165	153	- 1.4	1.96	- 8.1	65.61
170	159	+ 3.6	12.96	- 2.1	4.41
165	166	- 1.4	1.96	+ 4.9	24.01
164	163	- 2.4	5.76	+ 1.9	3.61
181	166	+ 14.6	213.16	+ 4.9	24.01
169	163	+ 2.6	6.76	+ 1.9	3.61
161	165	- 5.4	29.16	+ 3.9	15.21
162	167	- 4.4	19.36	+ 5.9	34.81
165	164	- 1.4	1.96	+ 2.9	8.41
163	162	- 3.4	11.56	+ 0.9	0.81
168	160	+ 1.6	2.56	- 1.1	1.21
169	159	+ 2.6	6.76	- 2.1	4.41
164	167	- 2.4	5.76	+ 5.9	34.81
180	157	+ 13.6	184.96	- 4.1	16.81
160	158	- 6.4	40.96	- 3.1	9.61
160		- 6.4	40.96		
167		+ 0.6	0.36		
174		+ 7.6	57.76		
168		+ 1.6	2.56		
165		- 1.4	1.96		
Σ 4160	3222		736.00		349.80

(a) Using the data presented in Table 7.2, the group means \bar{X}_1 and \bar{X}_2 are first computed.

$$\bar{X}_1 = \frac{\sum X_1}{n_1} = \frac{4160}{25} = 166.4 \text{ cm}; \quad \bar{X}_2 = \frac{\sum X_2}{n_2} = \frac{3222}{20} = 161.1 \text{ cm}.$$

(b) The deviations of the scores from the respective group means are worked out and squared, and the squared deviations of each group are totalled to give the respective sums of squares, $\sum(X_1 - \bar{X}_1)^2$ and $\sum(X_2 - \bar{X}_2)^2$.

$$\sum(X_1 - \bar{X}_1)^2 = 736.00 \text{ cm}^2; \quad \sum(X_2 - \bar{X}_2)^2 = 349.80 \text{ cm}^2.$$

(c) The pooled SD (\hat{s}) is computed using the sums of squares.

$$\hat{s} = \sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{736.00 + 349.80}{25 + 20 - 2}} = 5.025 \text{ cm}.$$

(d) The SE of the difference between means ($s_{\bar{X}_1 - \bar{X}_2}$) is computed using the pooled SD .

$$s_{\bar{X}_1 - \bar{X}_2} = \hat{s} \sqrt{\frac{n_1 + n_2}{n_1 n_2}} = 5.025 \sqrt{\frac{25 + 20}{25 \times 20}} = 1.508 \text{ cm}.$$

(e) The difference ($\bar{X}_1 - \bar{X}_2$) between the means is converted to t score.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{166.4 - 161.1}{1.508} = 3.515.$$

$$df = n_1 + n_2 - 2 = 25 + 20 - 2 = 43.$$

$$\begin{aligned} \text{[Alternatively, } t &= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2} \times \frac{n_1 + n_2}{n_1 n_2}}} \\ &= \frac{166.4 - 161.1}{\sqrt{\frac{736.00 + 349.80}{25 + 20 - 2} \times \frac{25 + 20}{25 \times 20}}} = 3.515.] \end{aligned}$$

(f) For different levels of significance, two-tail critical t scores ($df = 43$) are quoted below from the Table B of Appendix.

$$t_{.05(43)} = 2.017; \quad t_{.02(43)} = 2.416; \quad t_{.01(43)} = 2.695; \quad t_{.001(43)} = 3.532.$$

The computed t of 3.515 is found to be higher than the critical $t_{.01}$, but lower than the critical $t_{.001}$. So, the probability P of getting the observed difference between the means by chance due to random sampling amounts to less than 0.01, which may be considered too low. Hence, the H_0 may be rejected. So, the mean height of males differs significantly from that of females ($P < 0.01$).

Example 7.7.2.

The mean time for regeneration of amputated basal discs was found to be 18.6 hrs (SD 2.20) in 15 wild (nonmutated) animals of *Hydra vulgaris*, and 17.7 hrs (SD 2.70) in 12 somatic mutants of that species. Is the mean regeneration time significantly higher in the wild animals than in the mutants?

Solution :

The H_0 proposes that the mean regeneration time is not significantly higher in wild animals than in mutants. A *one-tail t test* is done to estimate the probability P of the H_0 being correct.

Wild animals : $\bar{X}_1 = 18.6$ hrs ; $s_1 = 2.20$ hrs ; $n_1 = 15$.

Mutants : $\bar{X}_2 = 17.7$ hrs ; $s_2 = 2.70$ hrs ; $n_2 = 12$.

(a) The pooled SD (\hat{s}) of the small unequal groups is first computed using their respective SDs (s_1 and s_2).

$$\hat{s} = \sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1 + n_2 - 2}} = \sqrt{\frac{(2.20)^2 \times (15-1) + (2.70)^2 \times (12-1)}{15+12-2}} = 2.433 \text{ hr.}$$

(b) The difference ($\bar{X}_1 - \bar{X}_2$) between the means is next converted to a t score.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\hat{s} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}} = \frac{18.6 - 17.7}{2.433 \sqrt{\frac{15+12}{15 \times 12}}} = 0.955.$$

$$df = n_1 + n_2 - 2 = 15 + 12 - 2 = 25.$$

(c) The computed t score is compared with *one-tail critical t scores* ($df = 25$) for different levels of significance (Table B of Appendix).

$$t_{.005(25)} = 2.787 ; \quad t_{.01(25)} = 2.485 ; \quad t_{.025(25)} = 2.060 ; \quad t_{.05(25)} = 1.708.$$

The computed t is found to be lower than the critical $t_{.05}$. So, the probability P of H_0 being correct is higher than 0.05 and may be considered too high. Hence, the H_0 cannot be rejected. So, the mean regeneration time is *not significantly higher* in wild animals ($P > 0.05$).

Example 7.7.3.

In a numerical operations test, the mean and SD of the scores of 16 boys were found to be 40.3 and 8.15 respectively ; these values amounted to 37.5 and 6.35, respectively, for 16 girls. Is there a significant difference between the means of boys and girls ?

Solution :

The H_0 proposes that there is no significant difference between the means of the two groups. To find the probability P of this H_0 being correct, a *two-tail t test* is done.

Boys : $\bar{X}_1 = 40.3$; $s_1 = 8.15$. Girls : $\bar{X}_2 = 37.5$; $s_2 = 6.35$. Size of each sample : $n = 16$.

(a) The difference ($\bar{X}_1 - \bar{X}_2$) between the means is converted to t score, using the unbiased SDs of the independent groups of equal size.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2 + s_2^2}{n}}} = \frac{40.3 - 37.5}{\sqrt{\frac{(8.15)^2 + (6.35)^2}{16}}} = 1.084.$$

$$df = 2(n - 1) = 2(16 - 1) = 30.$$

TESTING OF HYPOTHESIS

(b) The computed t score is compared with two-tail critical t scores ($df = 30$) for different levels of significance (Table B of Appendix).

$$t_{.05(30)} = 2.042 ; \quad t_{.02(30)} = 2.457 ; \quad t_{.01(30)} = 2.750.$$

As the computed t of 1.084 is lower than even the critical t for the 0.05 level of significance, the probability P of the H_0 being correct exceeds 0.05 and is considered too high. The H_0 cannot be rejected and it is inferred that the mean scores do not differ significantly ($P > 0.05$).

Example 7.7.4.

The body weights (kg) of 8 adult males and 8 adult females are presented in respectively the first and second columns of Table 7.3. Find whether or not the mean weight of males is significantly higher than that of females.

Solution :

The H_0 contends that the mean weight of males is not significantly higher than that of females. A one-tail t test is undertaken to estimate the probability P of this H_0 being correct.

Table 7.3. Table for computing mean body weights and sums of squares.

Weights (kg)		$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	$X_2 - \bar{X}_2$	$(X_2 - \bar{X}_2)^2$
Males (X_1)	Females (X_2)				
50	49	-7	49	-3	9
58	52	+1	1	0	0
60	51	+3	9	-1	1
55	56	-2	4	+4	16
59	55	+2	4	+3	9
56	53	-1	1	+1	1
54	52	-3	9	0	0
64	48	+7	49	-4	16
Σ 456	416		126		52

(a) Using the data presented in Table 7.3, the group means, \bar{X}_1 and \bar{X}_2 , are first computed. For each group, $n = 8$.

$$\bar{X}_1 = \frac{\sum X_1}{n} = \frac{456}{8} = 57 \text{ kg} ; \quad \bar{X}_2 = \frac{\sum X_2}{n} = \frac{416}{8} = 52 \text{ kg}.$$

(b) The deviations of scores of each group from its mean are worked out and squared. The squared deviations are totalled for each group to give the sum of squares. From Table 7.3,

$$\sum (X_1 - \bar{X}_1)^2 = 126 \text{ kg}^2 ; \quad \sum (X_2 - \bar{X}_2)^2 = 52 \text{ kg}^2.$$

(c) The sums of squares are used in computing the variances of the respective groups.

$$s_1^2 = \frac{\sum (X_1 - \bar{X}_1)^2}{n-1} = \frac{126}{8-1} = 18.0 \text{ kg}^2 ; \quad s_2^2 = \frac{\sum (X_2 - \bar{X}_2)^2}{n-1} = \frac{52}{8-1} = 7.4 \text{ kg}^2.$$

(d) The difference $(\bar{X}_1 - \bar{X}_2)$ between the means is converted to t score, using the variances.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2 + s_2^2}{n}}} = \frac{57 - 52}{\sqrt{\frac{18.0 + 7.4}{8}}} = 2.809 ; \quad df = 2(n - 1) = 2(8 - 1) = 14.$$

[Alternatively, the t score may be computed using the sums of squares, omitting step (c).

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sum(X_1 - \bar{X}_1)^2 + \sum(X_2 - \bar{X}_2)^2}{n(n-1)}}} = \frac{57 - 52}{\sqrt{\frac{126 + 52}{8(8-1)}}} = 2.809.]$$

(e) The computed t score is compared with *one-tail critical t scores* ($df = 14$) for different levels of significance (Table B of Appendix).

$$t_{.05(14)} = 1.761 ; \quad t_{.025(14)} = 2.145 ; \quad t_{.01(14)} = 2.624 ; \quad t_{.005(14)} = 2.977.$$

As the computed t of 2.809 is higher than the critical t for the 0.01 level of significance, the probability P of correctness of the H_0 is less than 0.01 and is, therefore, considered too low. So, the H_0 is rejected. It is inferred that the mean body weight of males is *significantly higher* than that of females ($P < 0.01$).

Example 7.7.5.

The mean and the SD of birthweights were found to be 2.9 kg and 0.65 kg respectively for 832 first-born infants, and 3.3 kg and 0.55 kg respectively for 608 third-born infants. Is the mean birthweight significantly higher in the third-born infants ?

Solution :

The H_0 proposes that the mean birthweight is not significantly higher in third-born infants. A *one-tail t test* is undertaken to find the probability P of this H_0 being correct.

For first-born infants : $\bar{X}_1 = 2.9$ kg ; $s_1 = 0.65$ kg ; $n_1 = 832$.

For third-born infants : $\bar{X}_2 = 3.3$ kg ; $s_2 = 0.55$ kg ; $n_2 = 608$.

(a) Because both the groups are large ($n_1 > 30$, $n_2 > 30$) and their sizes are not identical ($n_1 \neq n_2$), the t score for the difference ($\bar{X}_1 - \bar{X}_2$) between the means is computed using the respective SD s.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}} = \frac{2.9 - 3.3}{\sqrt{\frac{(0.65)^2}{608} + \frac{(0.55)^2}{832}}} = -12.295 ; \quad df = 832 + 608 - 2 = \infty.$$

(b) Because t distributions are bilaterally symmetrical, the negative sign of the computed t is ignored and the absolute value of the computed t score (viz., 12.295) is compared with the *one-tail critical t scores* ($df = \infty$) for different levels of significance, quoted from Table B of Appendix.

$$t_{.05(\infty)} = 1.645 ; \quad t_{.025(\infty)} = 1.960 ; \quad t_{.01(\infty)} = 2.326 ; \quad t_{.005(\infty)} = 2.576 ; \quad t_{.0005(\infty)} = 3.291.$$

The computed t being higher than even the critical t score for 0.0005 level of significance, the probability P of the H_0 being correct is lower than 0.0005. It is considered too low. So, the H_0 is rejected and it is inferred that the third-born infants have a *significantly higher* mean birthweight ($P < 0.0005$).

Example 7.7.6.

Find whether or not there is a significant difference between the mean winglength scores (mm) of the following two groups of houseflies sampled from two different habitats.

Winglengths (mm) :

- (a) Group 1 : 4.9, 5.2, 4.7, 5.3, 3.9, 5.4, 4.5, 4.9, 4.8, 5.0, 4.2, 4.8.
 (b) Group 2 : 3.1, 3.7, 3.6, 4.0, 3.3, 3.4, 3.3, 3.2, 3.0, 3.4.

Solution :

The H_0 proposes that there is no significant difference between the means. A two-tail t test is done, using the pooled SD (\hat{s}) of the two small independent groups of unequal sizes.

(a) Entering the data in the first two columns of Table 7.4, the group means, \bar{X}_1 and \bar{X}_2 , are computed.

$$n_1 = 12 ; \quad \bar{X}_1 = \frac{\sum X_1}{n_1} = \frac{57.6}{12} = 4.8 \text{ mm.} \quad n_2 = 10; \quad \bar{X}_2 = \frac{\sum X_2}{n_2} = \frac{34.0}{10} = 3.4 \text{ mm.}$$

Table 7.4. Table for computing mean winglengths and sums of squares.

Winglengths (mm)		$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	$X_2 - \bar{X}_2$	$(X_2 - \bar{X}_2)^2$
Group 1 (X_1)	Group 2 (X_2)				
4.9	3.1	+ 0.1	0.01	- 0.3	0.09
5.2	3.7	+ 0.4	0.16	+ 0.3	0.09
4.7	3.6	- 0.1	0.01	+ 0.2	0.04
5.3	4.0	+ 0.5	0.25	+ 0.6	0.36
3.9	3.3	- 0.9	0.81	- 0.1	0.01
5.4	3.4	+ 0.6	0.36	0	0
4.5	3.3	- 0.3	0.09	- 0.1	0.01
4.9	3.2	+ 0.1	0.01	- 0.2	0.04
4.8	3.0	0	0	- 0.4	0.16
5.0	3.4	+ 0.2	0.04	0	0
4.2		- 0.6	0.36		
4.8		0	0		
Σ 57.6	34.0		2.10		0.80

(b) The deviations of the scores of each group from its mean are worked out and squared. The squared deviations are totalled for each group to give the respective sums of squares which are then used in working out the pooled SD (\hat{s}). From Table 7.4,

$$\Sigma(X_1 - \bar{X}_1)^2 = 2.10 \text{ mm}^2 ; \quad \Sigma(X_2 - \bar{X}_2)^2 = 0.80 \text{ mm}^2 ;$$

$$\hat{s} = \sqrt{\frac{\Sigma(X_1 - \bar{X}_1)^2 + \Sigma(X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{2.10 + 0.80}{12 + 10 - 2}} = 0.38 \text{ mm.}$$

(c) The SE of the difference between means ($s_{\bar{X}_1 - \bar{X}_2}$) is computed using the pooled SD .

$$s_{\bar{X}_1 - \bar{X}_2} = \hat{s} \sqrt{\frac{n_1 + n_2}{n_1 n_2}} = 0.38 \sqrt{\frac{12 + 10}{12 \times 10}} = 0.163 \text{ mm.}$$

(d) The difference $(\bar{X}_1 - \bar{X}_2)$ between the means is converted to t score.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{4.8 - 3.4}{0.163} = 8.589 ; \quad df = n_1 + n_2 - 2 = 12 + 10 - 2 = 20.$$

(e) The computed t score is compared with the following *two-tail critical t scores* ($df = 20$) quoted from Table B of Appendix.

$$t_{.05(20)} = 2.086 ; \quad t_{.02(20)} = 2.528 ; \quad t_{.01(20)} = 2.845 ; \quad t_{.001(20)} = 3.850.$$

As the computed t of 8.589 is found to be higher than the critical $t_{.001}$, the probability P of correctness of the H_0 is less than 0.001 and may be considered too low. So, the H_0 is rejected, and it is inferred that there is a *significant difference* between the mean winglengths of the two groups ($P < 0.001$).

Example 7.7.7.

The mean and *SD* of steadiness test scores were found to be respectively 5.9 and 1.85 in a group of 40 women, and respectively 5.1 and 1.42 in a group of 42 men. Is the mean steadiness score significantly higher in women than in men ?

Solution :

The H_0 proposes that the mean steadiness score is not significantly higher in women than in men. A *one-tail t test* is undertaken to work out the probability P of this H_0 being correct.

$$\text{Women : } \bar{X}_1 = 5.9 ; \quad s_1 = 1.85 ; \quad n_1 = 40.$$

$$\text{Men : } \bar{X}_2 = 5.1 ; \quad s_2 = 1.42 ; \quad n_2 = 42.$$

(a) As both the groups are large (> 30) and of unequal sizes, the difference $(\bar{X}_1 - \bar{X}_2)$ between their means is computed using their *SDs*.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{5.9 - 5.1}{\sqrt{\frac{(1.85)^2}{42} + \frac{(1.42)^2}{40}}} = 2.203 ; \quad df = n_1 + n_2 - 2 = 40 + 42 - 2 = 80.$$

(b) The computed t is compared with the following *one-tail critical t scores* ($df = 80$) quoted from Table B of Appendix.

$$t_{.05(80)} = 1.664 ; \quad t_{.025(80)} = 1.990 ; \quad t_{.01(80)} = 2.374.$$

The computed t of 2.203 is found to be higher than the critical $t_{.025}$, but lower than the critical $t_{.01}$. So, the probability P of the H_0 being correct is less than 0.025, which may be considered too low. Hence, the H_0 is rejected, and it is inferred that the mean steadiness score is *significantly higher* in women than in men ($P < 0.025$).

2. t test for paired observations of small single groups by difference method

In a *single-group experiment*, the same group of individuals, randomly sampled from a population, serves first as the control group and subsequently as the experimental group. The

same group is thus treated with or exposed to successive levels of the independent variable ; after each such treatment or exposure, the dependent variable is measured in the group. The two sets of scores of the dependent variable, obtained from the single group after

its exposure to the respective levels of the independent variable, form *paired observations*, each individual having one pair of scores, and are *correlated* with each other. For example, the strength of the knee jerk reflex may be initially measured in a group of individuals after injecting them with placebo free from adrenaline (control group) ; they are then injected with adrenaline and the strength of the knee jerk reflex is subsequently measured again in them (experimental group). These two sets of knee jerk strength scores form paired and correlated observations. The effect of the independent variable on the dependent one is explored by estimating the significance of difference between the means of such paired observations by *t* test. However, the computation of *t* for such paired scores requires the use of the correlation coefficient (*r*) between the two sets of scores ; but *r* cannot be computed that correctly for small groups. So, to bypass the use of the correlation coefficient, a *t* test by the difference method is used for finding the significance of the difference between means of *paired observations of a small group* ($n < 30$) in such a single-group experiment.

For a *two-tail t test* by the difference method, the H_0 proposes that the mean difference (\bar{D}) between the paired scores does not differ significantly from 0. For a *one-tail t test*, the H_0 proposes that \bar{D} does not have a

significant positive (alternatively, negative) value. To estimate the probability *P* of correctness of the H_0 , the mean difference (\bar{D}) is converted to a *t* score. Where *n* is the number of pairs of scores, *D* is the difference between the scores of any pair, ΣD is the sum of all such differences between paired scores, \bar{D} and s_D are respectively the mean and the SD of those differences, and $s_{\bar{D}}$ is the SE of \bar{D} ,

$$\bar{D} = \frac{\Sigma D}{n} ; \quad s_D = \sqrt{\frac{\Sigma (D - \bar{D})^2}{n-1}} ; \quad s_{\bar{D}} = \frac{s_D}{\sqrt{n}} ;$$

$$t = \frac{\bar{D}}{s_{\bar{D}}} ; \quad df = n - 1.$$

Alternatively, *t* may be computed directly from the differences (*D*) between the paired scores and the squared values (D^2) of those differences.

$$t = \frac{\Sigma D}{\sqrt{\frac{n \Sigma D^2 - (\Sigma D)^2}{n-1}}} ; \quad df = n - 1.$$

The computed *t* score is then compared with the critical *t* scores with the same *df*. The difference between the paired observations is considered significant if the computed *t* equals or exceeds the critical *t* for the chosen level of significance ($P \leq \alpha$). On the contrary, the difference is not significant if the computed *t* is lower than the critical *t* for the chosen significance level ($P > \alpha$).

Example 7.7.8.

The achievement test scores of 10 students, before and after practice, are given below. Does practice make a significant difference in achievement test scores ?

Individuals	:	1	2	3	4	5	6	7	8	9	10
Achievement scores	:										
(i) before practice	:	72	67	90	97	84	92	65	75	80	69
(ii) after practice	:	120	81	110	103	109	137	115	82	110	89

Solution :

The H_0 contends that there is no significant difference between the paired scores, any observed difference being due to mere chances associated with random sampling. To estimate the probability P of correctness of this H_0 , a two-tail t test by the difference method is applicable as the paired scores belong to a small single group.

1. First method :

(a) The difference D is worked out between the paired scores (X_1 and X_2) of each individual and entered in Table 7.5. These differences for all the pairs of scores are totalled to give ΣD which is used in working out the mean difference \bar{D} .

$$n = 10 ; \quad \bar{D} = \frac{\Sigma D}{n} = \frac{265}{10} = 26.5.$$

Table 7.5. Table for t test of achievement scores by the difference method, using the mean difference.

Individuals	Achievement test scores		D ($X_2 - X_1$)	$D - \bar{D}$	$(D - \bar{D})^2$
	before practice (X_1)	after practice (X_2)			
1	72	120	+ 48	+ 21.5	462.25
2	67	81	+ 14	- 12.5	156.25
3	90	110	+ 20	- 6.5	42.25
4	97	103	+ 6	- 20.5	420.25
5	84	109	+ 25	- 1.5	2.25
6	92	137	+ 45	+ 18.5	342.25
7	65	115	+ 50	+ 23.5	552.25
8	75	82	+ 7	- 19.5	380.25
9	80	110	+ 30	+ 3.5	12.25
10	69	89	+ 20	- 6.5	42.25
Σ			265		2412.50

(b) The deviation of each D from \bar{D} is worked out and entered in Table 7.5. Each $(D - \bar{D})$ thus obtained is squared and the squared deviation $(D - \bar{D})^2$ is also entered in Table 7.5. The sum $\Sigma(D - \bar{D})^2$ of all the squared deviations is used in working out the SD of the differences (s_D), which is next used in computing the SE ($s_{\bar{D}}$) of \bar{D} .

$$s_D = \sqrt{\frac{\Sigma(D - \bar{D})^2}{n-1}} = \sqrt{\frac{2412.50}{10-1}} = 16.37 ; \quad s_{\bar{D}} = \frac{s_D}{\sqrt{n}} = \frac{16.37}{\sqrt{10}} = 5.177.$$

(c) \bar{D} is converted to t score using $s_{\bar{D}}$.

$$t = \frac{\bar{D}}{s_{\bar{D}}} = \frac{26.5}{5.177} = 5.119 ; \quad df = n - 1 = 10 - 1 = 9.$$

Table 7.6. Table for t test of achievement test scores by the difference method, using the squared differences.

Individuals	Achievement test scores		D ($X_2 - X_1$)	D^2
	before practice (X_1)	after practice (X_2)		
1	72	120	+ 48	2304
2	67	81	+ 14	196
3	90	110	+ 20	400
4	97	103	+ 6	36
5	84	109	+ 25	625
6	92	137	+ 45	2025
7	65	115	+ 50	2500
8	75	82	+ 7	49
9	80	110	+ 30	900
10	69	89	+ 20	400
Σ			+ 265	9435

2. Alternative method :

(a) The difference D is worked out between the paired scores (X_1 and X_2) of each individual and entered in Table 7.6. These differences for all the pairs of scores are totalled to give ΣD .

(b) Each such difference is squared to give D^2 . All such D^2 values are entered in Table 7.6 and totalled to give ΣD^2 .

(c) The t score is computed, using ΣD , ΣD^2 and the group size ($n = 10$).

$$t = \frac{\Sigma D}{\sqrt{\frac{n \Sigma D^2 - (\Sigma D)^2}{n-1}}} = \frac{265}{\sqrt{\frac{10 \times 9435 - (265)^2}{10-1}}} = 5.118 ; \quad df = n - 1 = 10 - 1 = 9.$$

The t score, computed by any of the two methods is then compared with the *two-tail critical t scores* ($df = 9$) for different levels of significance (Table B of Appendix).

$$t_{.05(9)} = 2.262 ; \quad t_{.02(9)} = 2.821 ; \quad t_{.01(9)} = 3.250 ; \quad t_{.001(9)} = 4.781.$$

As the computed t exceeds even the critical $t_{.001}$, the probability P of correctness of the H_0 is lower than 0.001 and is considered too low. So, the H_0 cannot be retained, and it is inferred that practice produces a significant difference in achievement test scores ($P < 0.001$).

Example 7.7.9.

The amounts of ethereal sulfates (mg) in 24 hours' urine of 8 individuals, kept successively on low and high protein diets, are given below. Is there a significant difference in the urinary excretion of ethereal sulfates due to changes in the dietary protein content ?

Individuals	:	1	2	3	4	5	6	7	8
Ethereal sulfates	:								
(i) low-protein diet	:	90	102	114	103	95	115	106	96
(ii) high-protein diet	:	123	125	138	115	132	141	126	121

Solution :

The H_0 proposes that there is no significant difference between the paired observations, any observed difference arising from mere chances of random sampling. A two-tail t test by difference method is applied to the paired observations of the small single group to find the probability P of the H_0 being correct.

1. First method :

Table 7.7. Table for t test of urinary ethereal sulfates by the difference method, using the mean difference.

Individuals	Ethereal sulfates		D ($X_2 - X_1$)	$D - \bar{D}$	$(D - \bar{D})^2$
	low-protein diet (X_1)	high-protein diet (X_2)			
1	90	123	+ 33	+ 8	64
2	102	125	+ 23	- 2	4
3	114	138	+ 24	- 1	1
4	103	115	+ 12	- 13	169
5	95	132	+ 37	+ 12	144
6	115	141	+ 26	+ 1	1
7	106	126	+ 20	- 5	25
8	96	121	+ 25	0	0
Total			+ 200		408

(a) The difference D is worked out between the paired scores (X_1 and X_2) of each individual and entered in Table 7.7. These differences for all the pairs of scores are totalled to give ΣD which is used in working out the mean difference \bar{D} .

$$n = 8 ; \quad \bar{D} = \frac{\Sigma D}{n} = \frac{200}{8} = 25 \text{ mg.}$$

(b) The deviation of each D from \bar{D} is worked out and entered in Table 7.7. Each $(D - \bar{D})$ thus obtained is squared and the squared deviation $(D - \bar{D})^2$ is also entered in that table. Then, the sum $\Sigma(D - \bar{D})^2$ of all the squared deviations is used in computing the SD of the differences (s_D) which is used in turn in computing the SE ($s_{\bar{D}}$) of \bar{D} .

$$s_D = \sqrt{\frac{\Sigma(D - \bar{D})^2}{n-1}} = \sqrt{\frac{408}{8-1}} = 7.63 \text{ mg ;} \quad s_{\bar{D}} = \frac{s_D}{\sqrt{n}} = \frac{7.63}{\sqrt{8}} = 2.70 \text{ mg.}$$

(c) \bar{D} is then converted to t score using $s_{\bar{D}}$.

$$t = \frac{\bar{D}}{s_{\bar{D}}} = \frac{25}{2.70} = 9.259 ; \quad df = n - 1 = 8 - 1 = 7.$$

2. Alternative method :

(a) The difference D is worked out between the paired scores (X_1 and X_2) of each individual and entered in Table 7.8. These differences are totalled for all the pairs to give ΣD .

(b) Each such difference D is squared to D^2 . All the D^2 values are entered in Table 7.8 and totalled to give ΣD^2 .

Table 7.8. Table for t test of ethereal sulfates by the difference method, using the squared differences.

Individuals	X_1	X_2	$D = (X_2 - X_1)$	D^2
1	90	123	+ 33	1089
2	102	125	+ 23	529
3	114	138	+ 24	576
4	103	115	+ 12	144
5	95	132	+ 37	1369
6	115	141	+ 26	676
7	106	126	+ 20	400
8	96	121	+ 25	625
Total			+ 200	5408

(c) The t score is computed using ΣD , ΣD^2 and $n (= 8)$.

$$t = \frac{\Sigma D}{\sqrt{\frac{n \Sigma D^2 - (\Sigma D)^2}{n-1}}} = \frac{200}{\sqrt{\frac{8 \times 5408 - (200)^2}{8-1}}} = 9.262 ; \quad df = n - 1 = 8 - 1 = 7.$$

Interpretation :

The t score, computed by either of the two methods, is next compared with the two-tail critical t scores ($df = 7$) for different levels of significance (Table B of Appendix).

$$t_{.05(7)} = 2.365 ; \quad t_{.02(7)} = 2.998 ; \quad t_{.01(7)} = 3.499 ; \quad t_{.001(7)} = 5.405.$$

As the computed t far exceeds the critical $t_{.001}$, the probability P of correctness of H_0 is less than 0.001 and is considered too low. So, the H_0 is rejected, and it is inferred that changes in dietary protein contents produce a significant difference in the urinary ethereal sulfate excretion ($P < 0.001$).

Example 7.7.10.

Hourly oxygen consumptions (in ml per 100 g of bodyweight) were found to be as follows in a sample of nine parakeets, respectively before and after exposure to a pesticide.

Individual	:	1	2	3	4	5	6	7	8	9
Before pesticide (X_1)	:	182	157	173	185	175	168	179	159	170
After pesticide (X_2)	:	157	122	140	166	138	148	148	130	147

Find whether or not the oxygen consumption is significantly higher before exposure to the pesticide than that after exposure to it.

Solution :

The H_0 contends that the oxygen consumption is not significantly higher before exposure than after it. A one-tail t test is undertaken by the difference method to work out the probability P of correctness of this H_0 as the paired scores belong to a small single group.

1. First method :

(a) The difference D is worked out between the paired scores (X_1 and X_2) of each individual and entered in Table 7.9. All these D values are totalled to give ΣD which is used in computing the mean difference \bar{D} .

$$n = 9 ; \quad \bar{D} = \frac{\Sigma D}{n} = \frac{252}{9} = 28.0 \text{ ml.}$$

Table 7.9. Table for t test of oxygen consumptions by the difference method, using the mean difference.

Individuals	Oxygen consumptions		D ($X_1 - X_2$)	$D - \bar{D}$	$(D - \bar{D})^2$
	before pesticide (X_1)	after pesticide (X_2)			
1	182	157	+ 25	- 3	9
2	157	122	+ 35	+ 7	49
3	173	140	+ 33	+ 5	25
4	185	166	+ 19	- 9	81
5	175	138	+ 37	+ 9	81
6	168	148	+ 20	- 8	64
7	179	148	+ 31	+ 3	9
8	159	130	+ 29	+ 1	1
9	170	147	+ 23	- 5	25
Σ			252		344

(b) The deviation of each D value from \bar{D} is worked out and entered in Table 7.9. Each such $(D - \bar{D})$ is squared and the squared deviation $(D - \bar{D})^2$ is also entered in the table. The sum $\Sigma(D - \bar{D})^2$ of all the squared deviations is used in working out the SD of the differences (s_D), which is used in turn in computing the SE ($s_{\bar{D}}$) of \bar{D} .

$$s_D = \sqrt{\frac{\Sigma(D - \bar{D})^2}{n-1}} = \sqrt{\frac{344}{9-1}} = 6.557 \text{ ml ;} \quad s_{\bar{D}} = \frac{s_D}{\sqrt{n}} = \frac{6.557}{\sqrt{9}} = 2.186 \text{ ml.}$$

(c) \bar{D} is converted to t score, using $s_{\bar{D}}$.

$$t = \frac{\bar{D}}{s_{\bar{D}}} = \frac{28.0}{2.186} = 12.809 ; \quad df = n - 1 = 9 - 1 = 8.$$

2. Alternative method :

(a) The difference D is worked out between the paired scores (X_1 and X_2) of each individual and entered in Table 7.10. All these D values are totalled to give ΣD .

(b) Each D is squared to D^2 . All the D^2 values are entered in Table 7.10 and totalled to give ΣD^2 .

Table 7.10. Table for t test of oxygen consumptions by the difference method, using the squared differences.

Individuals	X_1	X_2	$D = (X_1 - X_2)$	D^2
1	182	157	+ 25	625
2	157	122	+ 35	1225
3	173	140	+ 33	1089
4	185	166	+ 19	361
5	175	138	+ 37	1369
6	168	148	+ 20	400
7	179	148	+ 31	961
8	159	130	+ 29	841
9	170	147	+ 23	529
Σ			252	7400

(c) The t score is computed, using ΣD and ΣD^2 .

$$t = \frac{\Sigma D}{\sqrt{\frac{n \Sigma D^2 - (\Sigma D)^2}{n-1}}} = \frac{252}{\sqrt{\frac{9 \times 7400 - (252)^2}{9-1}}} = 12.810 ; \quad df = n - 1 = 9 - 1 = 8.$$

Interpretation :

The t score, computed by either of the two methods, is next compared with the *one-tail critical t scores* ($df = 8$) for different levels of significance (Table B of Appendix).

$$t_{.05(8)} = 1.860 ; \quad t_{.025(8)} = 2.306 ; \quad t_{.01(8)} = 2.896 ; \quad t_{.005(8)} = 3.355 ; \quad t_{.0005(8)} = 5.041.$$

As the computed t far exceeds the *critical 1-tail $t_{.0005}$* , the probability P of the H_0 being correct is less than 0.0005 and is considered too low. So, the H_0 is rejected, and it is inferred that the mean oxygen consumption is *significantly higher* before exposure to the pesticide than that after the exposure ($P < 0.0005$).

3. t test for paired observations of matched-pair and large single groups

Two *equivalent or matched-pair groups* may be constituted by including in one group such individuals, each of whom is matched with an individual of the other group with respect to an initially measured variable, either identical with or related to the dependent variable to be studied subsequently. The two matched groups are then exposed to or treated with *different levels* of the independent variable — one of the groups may, for example, serve as the *control group* treated with placebo devoid of the independent variable (*level 1* of treatment)

while the other serves as the *experimental group* exposed to the *level 2* of treatment consisting of a given dose of the independent variable. The dependent variable is subsequently measured in the individuals of both the groups. The two sets of final scores of the dependent variable constitute *paired and correlated observations* (page 126). For a t test of the difference between means of two such equivalent groups, the product-moment correlation coefficient (Pearson's r , vide § 8.2) between the two sets of scores has to be used in computing the $SE (s_{\bar{X}_1 - \bar{X}_2})$ of the difference.

For a *single-group experiment using a large group* ($n \geq 30$), the t test by the difference method, proposed for a small single-group experiment (pages 126-127), is not appropriate. Instead, the correlation coefficient r (vide § 8.2) computed between the paired scores has to be used here also in working out $s_{\bar{X}_1 - \bar{X}_2}$.

Thus, for n pairs of scores of either *matched-pair groups* or *large single groups*,

$$s_1 = \sqrt{\frac{\sum (X_1 - \bar{X}_1)^2}{n-1}} ; \quad s_2 = \sqrt{\frac{\sum (X_2 - \bar{X}_2)^2}{n-1}} ;$$

$$s_{\bar{X}_1} = \frac{s_1}{\sqrt{n}} ; \quad s_{\bar{X}_2} = \frac{s_2}{\sqrt{n}} ;$$

$$r = \frac{\sum (X_1 - \bar{X}_1)(X_2 - \bar{X}_2)}{\sqrt{\sum (X_1 - \bar{X}_1)^2 \sum (X_2 - \bar{X}_2)^2}} ;$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{(s_{\bar{X}_1})^2 + (s_{\bar{X}_2})^2 - 2r \cdot s_{\bar{X}_1} \cdot s_{\bar{X}_2}} ;$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} ; \quad df = n - 1.$$

To find the P of the H_0 being correct, the computed t is compared with critical t scores, with the same df , but for different levels of significance. The inference is drawn according as the P is considered too low or too high, as in the preceding examples of t test.

Example 7.7.11.

The strengths of kneejerk reflexes (in degrees of arc) as measured in a group of 30 men under tensed and relaxed conditions, respectively, are given in the first three columns of Table 7.11. Find whether or not the mean kneejerk strength in the tensed condition is significantly different from that in the relaxed condition.

Solution :

The H_0 contends that the mean kneejerk strengths are not significantly different in the two conditions. Because it is a *large single-group experiment*, a *two-tail t test* is undertaken, using the correlation coefficient r between the paired scores of the two sets, to work out the probability P of the H_0 being correct.

(a) The sum of the scores of each set are used to work out the mean of that set.

$$\bar{X}_1 = \frac{\sum X_1}{n} = \frac{900}{30} = 30 ; \quad \bar{X}_2 = \frac{\sum X_2}{n} = \frac{720}{30} = 24.$$

(b) \bar{X}_1 and \bar{X}_2 are then used in computing respectively $(X_1 - \bar{X}_1)$ and $(X_2 - \bar{X}_2)$ values, each of which is squared to give respectively $(X_1 - \bar{X}_1)^2$ and $(X_2 - \bar{X}_2)^2$ (Table 7.11). The respective *sums of squares*, viz., $\sum (X_1 - \bar{X}_1)^2$ and $\sum (X_2 - \bar{X}_2)^2$, are then worked out (Table 7.11) and used in computing the *SDs* (s_1 and s_2) of the two sets of scores.

$$s_1 = \sqrt{\frac{\sum (X_1 - \bar{X}_1)^2}{n-1}} = \sqrt{\frac{462}{30-1}} = 3.99 ; \quad s_2 = \sqrt{\frac{\sum (X_2 - \bar{X}_2)^2}{n-1}} = \sqrt{\frac{594}{30-1}} = 4.53.$$

(c) The *SEs* of the two means, viz., $s_{\bar{X}_1}$ and $s_{\bar{X}_2}$, are computed, using s_1 and s_2 respectively.

$$s_{\bar{X}_1} = \frac{s_1}{\sqrt{n}} = \frac{3.99}{\sqrt{30}} = 0.728 ; \quad s_{\bar{X}_2} = \frac{s_2}{\sqrt{n}} = \frac{4.53}{\sqrt{30}} = 0.827.$$

(d) The product of $(X_1 - \bar{X}_1)$ and $(X_2 - \bar{X}_2)$ is worked out for each individual, and these products for all the individuals are totalled to give the *sum of products*, viz., $\sum (X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$. (See Table 7.11)

Table 7.11. Table for computing the SDs and the correlation coefficient of kneejerk scores.

Individuals	Kneejerk strengths		$X_1 - \bar{X}_1$	$(X_1 - \bar{X}_1)^2$	$X_2 - \bar{X}_2$	$(X_2 - \bar{X}_2)^2$	$(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$
	tensed (X_1)	relaxed (X_2)					
1	30	25	0	0	+ 1	1	0
2	25	27	- 5	25	+ 3	9	- 15
3	31	28	+ 1	1	+ 4	16	+ 4
4	24	20	- 6	36	- 4	16	+ 24
5	31	22	+ 1	1	- 2	4	- 2
6	28	16	- 2	4	- 8	64	+ 16
7	27	31	- 3	9	+ 7	49	- 21
8	34	27	+ 4	16	+ 3	9	+ 12
9	36	23	+ 6	36	- 1	1	- 6
10	35	26	+ 5	25	+ 2	4	+ 10
11	37	29	+ 7	49	+ 5	25	+ 35
12	30	32	0	0	+ 8	64	0
13	32	27	+ 2	4	+ 3	9	+ 6
14	28	15	- 2	4	- 9	81	+ 18
15	26	14	- 4	16	- 10	100	+ 40
16	29	20	- 1	1	- 4	16	+ 4
17	24	20	- 6	36	- 4	16	+ 24
18	25	30	- 5	25	+ 6	36	- 30
19	31	22	+ 1	1	- 2	4	- 2
20	34	25	+ 4	16	+ 1	1	+ 4
21	35	27	+ 5	25	+ 3	9	+ 15
22	28	20	- 2	4	- 4	16	+ 8
23	24	27	- 6	36	+ 3	9	- 18
24	31	23	+ 1	1	- 1	1	- 1
25	35	27	+ 5	25	+ 3	9	+ 15
26	30	24	0	0	0	0	0
27	36	26	+ 6	36	+ 2	4	+ 12
28	31	25	+ 1	1	+ 1	1	+ 1
29	28	22	- 2	4	- 2	4	+ 4
30	25	20	- 5	25	- 4	16	+ 20
Σ	900	720		462		594	177

(e) The correlation coefficient (r) is computed between the paired scores of the two sets, using the sum of products and the sums of squares.

$$r = \frac{\Sigma(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)}{\sqrt{\Sigma(X_1 - \bar{X}_1)^2 \Sigma(X_2 - \bar{X}_2)^2}} = \frac{177}{\sqrt{462 \times 594}} = + 0.34.$$

(f) The SE of the difference between the means, $s_{\bar{X}_1 - \bar{X}_2}$, is worked out using the computed $s_{\bar{X}_1}$, $s_{\bar{X}_2}$ and r values.

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{(s_{\bar{X}_1})^2 + (s_{\bar{X}_2})^2 - 2r \cdot s_{\bar{X}_1} \cdot s_{\bar{X}_2}} = \sqrt{(0.728)^2 + (0.827)^2 - 2 \times 0.34 \times 0.728 \times 0.827} = 0.897.$$

(g) The difference between the means is then converted to t score.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{30 - 24}{0.897} = 6.689 ; \quad df = n - 1 = 30 - 1 = 29.$$

Interpretation :

The computed t is compared with the *two-tail critical t scores* ($df = 29$) for different levels of significance (Table B of Appendix).

$$t_{.05(29)} = 2.045 ; \quad t_{.02(29)} = 2.462 ; \quad t_{.01(29)} = 2.756 ; \quad t_{.001(29)} = 3.659.$$

As the computed t exceeds the *critical $t_{.001}$* , the probability P of the H_0 being correct is less than 0.001 and is considered too low. So, the H_0 is rejected and it is inferred that the mean kneejerk in the tensed condition is *significantly different* from that in the relaxed condition ($P < 0.001$).

Example 7.7.12.

The scores, obtained by 40 subjects in two successive trials of a sensory-motor test, had the following respective means and standard deviations.

$$\begin{aligned} \text{Trial 1 : } & \bar{X}_1 = 48.8 ; \quad s_1 = 9.87 ; \\ \text{Trial 2 : } & \bar{X}_2 = 55.7 ; \quad s_2 = 11.55. \quad n = 40. \end{aligned}$$

The correlation coefficient (r) between the paired scores of the two trials worked out to be +0.58.

Find whether or not the mean score of trial 2 is significantly higher than that of trial 1.

Solution :

The H_0 contends that the mean score of trial 2 is not significantly higher than that of trial 1. To estimate the probability P of correctness of this H_0 , a *one-tail t test* is undertaken using the correlation coefficient (r) between the paired scores belonging to a *large single group*.

(a) The standard errors, $s_{\bar{X}_1}$ and $s_{\bar{X}_2}$, of the respective means are computed, using the respective standard deviations, and are in turn used in working out the *SE of difference* ($s_{\bar{X}_1 - \bar{X}_2}$) between the means.

$$s_{\bar{X}_1} = \frac{s_1}{\sqrt{n}} = \frac{9.87}{\sqrt{40}} = 1.56 ; \quad s_{\bar{X}_2} = \frac{s_2}{\sqrt{n}} = \frac{11.55}{\sqrt{40}} = 1.83 ; \quad r = +0.58.$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{(s_{\bar{X}_1})^2 + (s_{\bar{X}_2})^2 - 2r \cdot s_{\bar{X}_1} \cdot s_{\bar{X}_2}} = \sqrt{(1.56)^2 + (1.83)^2 - 2 \times 0.58 \times 1.56 \times 1.83} = 1.572.$$

(b) The difference ($\bar{X}_2 - \bar{X}_1$) between the means is converted to t score.

$$t = \frac{\bar{X}_2 - \bar{X}_1}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{55.7 - 48.8}{1.572} = 4.389 ; \quad df = n - 1 = 40 - 1 = 39.$$

The computed t is compared with the *one-tail critical t scores* ($df = 39$) for different levels of

significance.

$$t_{.05(39)} = 1.6885 ; \quad t_{.01(39)} = 2.426 ; \quad t_{.005(39)} = 2.708 ; \quad t_{.0005(39)} = 3.558.$$

As the computed t exceeds the critical $t_{.0005}$, the probability P of the H_0 being correct is less than 0.0005 and is considered too low. So, the H_0 is rejected, and it is inferred that the mean score of trial 2 is significantly higher than that of trial 1 ($P < 0.0005$).

GLOSSARY

- alternative hypothesis** : the hypothesis which proposes that the result of the experiment is significant.
- equivalent-group experiment** : an experiment in which groups to be exposed to different levels of the independent variable are initially matched with respect to a variable either identical with or related to the dependent variable.
- errors of inference** : errors in either accepting or rejecting the null hypothesis depending on probabilities.
- homoscedasticity** : the assumption that the groups, drawn for an experiment, initially have homogeneous variances differing only due to sampling errors.
- independent group experiment** : an experiment in which groups to be exposed to different levels of the independent variable consist of such separate sets of individuals as are not related to each other.
- level of significance** : the highest level of probability upto which the probability of correctness of the null hypothesis (H_0) is considered so low that the H_0 is rejected.
- null hypothesis** : the hypothesis which contests the alternative hypothesis and proposes that the result of the experiment is not significant, being merely due to the chance use of a particular random sample.
- one-tail test** : a test for finding whether or not the mean of one group is significantly higher (alternatively, lower) than that of another in an experiment.
- paired observations** : observations of a single-group or equivalent-group experiment where each score of one set of observations is paired with a score of another set, giving rise to a correlation between the two sets of scores.
- single-group experiment** : an experiment in which the same group of individuals is exposed successively to different levels of the independent variable.
- two-tail test** : a test for finding whether or not there is a significant difference between the means of two groups in an experiment, irrespective of which of them is higher or lower than the other.
- type I error** : error of inference owing to the wrongful rejection of a correct null hypothesis, depending on the level of significance.
- type II error** : error of inference owing to the wrongful acceptance of a wrong null hypothesis.

8. CORRELATION AND REGRESSION

Correlation coefficients measure quantitatively the relationship between variables. *Regression* predicts the most likely value of a variable from the value(s) of one or more other variables.

Bivariate statistics analyze the data of two variables in a sample, measure the relationship between two variables, or predict the value of one variable from the given value of another variable in the same individual. *Multivariate statistics* analyze the data of more than two variables, measure their relationships, or predict the most likely value of one variable from the given values of two or more others.

8.1 CORRELATION

Correlation explores the *magnitude and direction of association* between two or more variables, i.e., how far the variations of a variable are related to those of one or more other variables in the same individual. It thus gives the magnitude and the algebraic sign of concomitant variations of variables.

Types

(a) Correlation is either *linear* or *nonlinear*, according as the relation between the variables can be described by a straight or a curved line. In linear correlation, the magnitude of change of one variable bears a constant ratio to that of the other variable(s).

(b) Correlation may be *positive* if the high and low magnitudes of one variable are associated with respectively the high and low magnitudes of the other ; it is *negative* if the high magnitude of one is associated with the low magnitude of the other.

(c) Both linear and nonlinear correlations

may be *simple* or *multiple*, according as they measure the relation either between two variables or between one variable and the weighted sum of two or more others.

For example, the correlation between body surface area and the weighted sum of height and weight is a multiple correlation. The correlation between the total RBC count and the blood haemoglobin concentration is simple and linear, while that between the substrate concentration and the initial velocity of enzyme action is simple but nonlinear, either hyperbolic or sigmoid.

Properties

1. Correlation holds good only within the limits of the population and other conditions in which it is estimated. It *cannot be generalized beyond those limits*.

2. It *may not indicate any cause and effect relationship* between the variables. It merely indicates an association between their changes without inferring whether or not the change in one has caused the change in the other. It is possible that the correlation has resulted from the common influence of some other variable(s) on the correlated variables.

3. Correlation *cannot predict* the value of one variable from that of another.

4. Correlation holds good even if the variables involved are free to *vary at random*, not being "fixed" or deliberately controlled by the investigator.

5. Correlation suffers from *sampling errors*. So, correlation coefficients of samples drawn from a population form a *sampling distribution* around the parametric correlation coefficient as the mean. The *SD* of such a sampling

distribution is the *SE* of the correlation coefficient, which is used in testing the significance of the computed correlation coefficient.

8.2 PRODUCT-MOMENT CORRELATION

The product-moment correlation coefficient (r), formulated by the British mathematician Karl Pearson (1857-1936), is a measure of the magnitude and direction (algebraic sign) of the relation between two variables in a sample when their relationship can be described by a straight line. It is thus a coefficient of *simple linear correlation*. Where X and Y are the scores of the variables, \bar{X} and \bar{Y} are their respective means, s_X and s_Y are the respective standard deviations, z_X and z_Y are the standard z scores of X and Y respectively, and n is the number of individuals (or pairs of scores) in the sample,

$$z_X = \frac{X - \bar{X}}{s_X}; \quad z_Y = \frac{Y - \bar{Y}}{s_Y};$$

$$r = \frac{\sum z_X z_Y}{n} = \frac{\sum \left(\frac{X - \bar{X}}{s_X} \times \frac{Y - \bar{Y}}{s_Y} \right)}{n};$$

$$\therefore r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{ns_X s_Y}.$$

This is the basic formula for r .

Assumptions

To compute and use the Pearson's r , it should be reasonable to assume that :

(a) both the variables are *continuous measurement variables* of either interval or ratio type, r being inapplicable to nominal, discontinuous and ordinal variables ;

(b) the variables have *unimodal and fairly symmetrical distributions* in the population without marked skewness, although they need not be normally distributed ;

(c) the paired scores of the two variables in each individual are *independent* of such paired scores of all other individuals in the sample ;

(d) there exists a *linear relationship* between the scores of the variables. In other words, the *scattergram* of their scores fits with a linear model (page 34).

Properties

1. The magnitude and the algebraic sign of r indicate respectively *the degree and the direction of linear relationship* between the variables. The value of r ranges between -1.00 and $+1.00$, and does not bear the unit of any of the variables. A value of 0 indicates the absence of any linear correlation. The closer is the value of r to -1.00 or $+1.00$, the stronger or higher is the relation between the variables.

2. Even if a constant number is added to or subtracted from all the scores of the two variables, or if they are divided or multiplied by a constant number, Pearson's r remains the same.

3. A *positive* r indicates that an individual, having a high score in one variable, is likely to possess a high score in the other variable ; on the other hand, an individual with a low score in one is expected to have a low score in the other too. A perfect positive correlation is indicated if r amounts to $+1.00$. In such a case, each individual has the same z_X and z_Y scores for his X and Y scores.

$$z_X = z_Y ;$$

$$\text{or, } X = \bar{X} + (Y - \bar{Y}) s_X / s_Y, \quad \text{and}$$

$$Y = \bar{Y} + (X - \bar{X}) s_Y / s_X.$$

A *negative* r indicates, on the contrary, that a high score of one variable is likely to be associated with a low score of the other, and vice versa. For a perfect negative correlation, r amounts to -1.00 . For each individual in such a case,

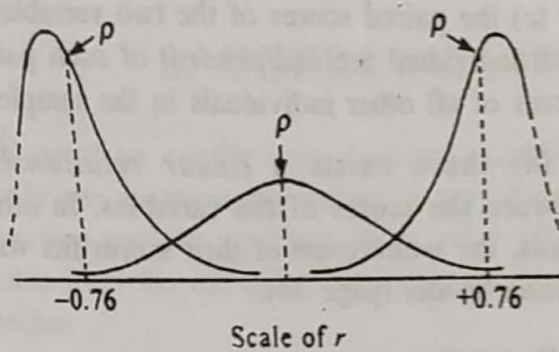


Fig. 8.1. Three sampling distributions of r .

$$z_X = -z_Y;$$

$$\text{or, } X = \bar{X} - (Y - \bar{Y}) s_X/s_Y, \text{ and}$$

$$Y = \bar{Y} - (X - \bar{X}) s_Y/s_X.$$

Seldom, however, a perfect correlation like +1.00 or -1.00 is obtained in practice. In many cases, a correlation of 0.60 is considered quite high — even a much lower correlation like 0.30 may be useful if the probability (P) of its random occurrence is too low.

4. Due to random variations of the variables in samples, r suffers from *sampling errors* and varies from sample to sample drawn from the same population. So, sample r values form a *sampling distribution* around the parametric correlation coefficient ρ . The sampling distribution is nearly symmetrical if ρ amounts to zero, but is progressively skewed positively or negatively as ρ approaches -1.00 and +1.00 respectively (Fig. 8.1). The SD of this sampling distribution is the *SE* (s_r) of the sample r and is used in finding the significance of the computed r .

5. Pearson's r (or any other correlation coefficient) does not directly give the percentage relationship or dependence of one variable on another, unless its value is +1.00, -1.00 or 0.00. In many cases, however, the squared value of the correlation coefficient gives the *percentage dependence* of the variables and is called the *coefficient of determination*. Thus, if r amounts to +0.60, only a 0.36 proportion or 36% of the total

variance of one variable may be associated with the variance of the other; the remaining proportion, 0.64 here, is the *coefficient of nondetermination* indicating such proportion of the variance of one variable as is not associated with that of the other. In some cases, however, the percentage dependence has a value different from r^2 .

Computations from ungrouped data

1. Using unbiased SD :

It may be recalled that for ungrouped data of small samples, unbiased standard deviations (s_X and s_Y) of variables X and Y are computed using the degrees of freedom, viz., $(n - 1)$, instead of the sample size (page 52). For such samples, Pearson's r is computed between the variables X and Y by the modified basic formula using their respective unbiased SDs (s_X and s_Y), the sample size (n), and $\Sigma(X - \bar{X})(Y - \bar{Y})$, the *sum of products*. (Example 8.2.1.)

$$s_X = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}}; \quad s_Y = \sqrt{\frac{\Sigma(Y - \bar{Y})^2}{n - 1}};$$

$$r = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{(n - 1) s_X s_Y}.$$

2. Using covariance :

Covariance is the 'conjoint variance', a measure of the associated variations of the scores of two variables. It is given by the arithmetic mean of the products of deviations of raw scores of those variables from their respective means.

$$\text{Cov}(X, Y) = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{n}, \text{ or,}$$

$$\text{Cov}(X, Y) = \frac{\Sigma XY}{n} - \frac{\Sigma X \Sigma Y}{n^2}.$$

Where s_X and s_Y are the unbiased standard deviations computed using $(n - 1)$ in the denominator,

$$\text{Cov}(X, Y) = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{n-1}, \text{ and}$$

$$r = \frac{\text{Cov}(X, Y)}{s_X s_Y}. \text{ (See Examples 8.2.1. to 8.2.3.)}$$

3. Using the sums of squares :

The product-moment r can also be computed using the sum of products as the numerator and the sums of squares, viz., $\sum(X - \bar{X})^2$ and $\sum(Y - \bar{Y})^2$, in the denominator.

$$s_X = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}}; \quad s_Y = \sqrt{\frac{\sum(Y - \bar{Y})^2}{n-1}};$$

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{(n-1)s_X s_Y}$$

$$\text{or, } r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{(n-1)\sqrt{\frac{\sum(X - \bar{X})^2}{n-1} \times \frac{\sum(Y - \bar{Y})^2}{n-1}}}$$

$$= \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}}.$$

(See Example 8.2.3.)

4. Using the raw scores :

Covariance as well as s_X and s_Y (page 52) may also be directly worked out from the raw scores (X and Y) of the variables. So, r can also be worked out directly from the raw scores. (See Example 8.2.2.)

$$\text{Cov}(X, Y) = \frac{\sum XY}{n} - \frac{\sum X \sum Y}{n^2};$$

$$s_X = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}; \quad s_Y = \sqrt{\frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n}\right)^2};$$

$$r = \frac{\text{Cov}(X, Y)}{s_X s_Y}$$

$$= \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}.$$

Significance of r

1. Test of H_0 that ρ is zero :

In this case, the H_0 contends that the

parametric (population) correlation coefficient (ρ) amounts to zero, and the observed deviation of the computed r from 0 has resulted from mere chances of random sampling. Because r has a bilaterally symmetrical sampling distribution when ρ is zero (Fig. 8.1), the computed r can be transformed into the t score for interpretation, using the standard error (s_r) of r . The df of the computed t should be taken as $n-2$, because any two pairs of raw scores lose their freedom for change in order to keep \bar{X} and \bar{Y} constant as estimates of the respective population means in the computation of r .

$$s_r = \sqrt{\frac{1-r^2}{n-2}}; \quad t = \frac{r}{s_r}; \quad df = n-2.$$

The computed t score is compared for a two-tail test with the two-tail critical t scores, with the same df , but for different levels of significance. The computed r is considered significant at or below that level of significance whose critical t_α either equals or is lower than the computed t ($P \leq \alpha$). But if the computed t is lower than the critical t score for either the chosen α or the α of 0.05, the H_0 cannot be rejected and the computed r is not significant.

2. Test of H_0 that ρ has a given value other than 0 — Fisher's z transformation :

The H_0 contends here that ρ has a specified value — other than zero — and that any observed deviation of the computed r from this given value has resulted from chances of random sampling. Student's t test cannot be used here because r has asymmetric and

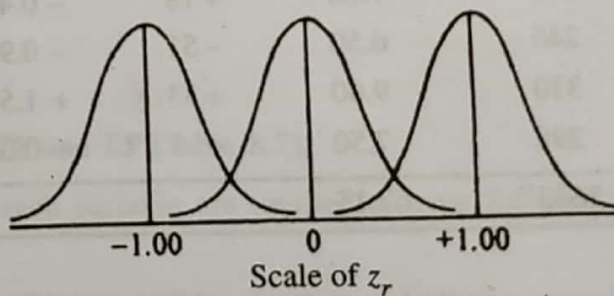


Fig. 8.2. Three sampling distributions of z_r .

skewed sampling distributions if $\rho \neq 0$. In such cases, the computed r is logarithmically transformed into Fisher's z_r , which has bilaterally symmetrical and approximately normal sampling distributions around its parameter ζ irrespective of the values of ρ and ζ (Fig. 8.2). So, the computed r and the given ρ are logarithmically transformed into z_r and ζ , respectively. The deviation of z_r from ζ is converted to the standard z score which is referred to the unit normal curve areas for

working out the probability (P) of the H_0 being correct.

$$z_r = 1.1513 \log \left[\frac{1+r}{1-r} \right]; \quad s_{z_r} = \frac{1}{\sqrt{n-3}};$$

$$\zeta = 1.1513 \log \left[\frac{1+\rho}{1-\rho} \right];$$

$$z = \frac{z_r - \zeta}{s_{z_r}} = (z_r - \zeta) \sqrt{n-3};$$

$$P = 2 [0.5000 - (\text{area of unit normal curve from its } \mu \text{ to the computed } z)].$$

Example 8.2.1.

Find whether or not there is a significant correlation between O_2 consumption (ml per minute) and pulmonary minute ventilation (litres per minute) using the following data.

Individuals	:	1	2	3	4	5	6	7	8	9
O_2 consumption	:	281	246	369	330	258	315	246	330	298
Ventilation	:	6.55	7.10	9.00	8.50	6.00	7.00	6.50	9.00	7.50

Solution :

1. First method :

(a) The scores of O_2 consumption (X) and ventilation (Y) are used to compute their means (\bar{X} and \bar{Y}) and unbiased SDs (s_X and s_Y) using Table 8.1.

Table 8.1. Table for computing r using unbiased SDs.

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
281	6.55	-16	-0.91	256	0.8281	+ 14.56
246	7.10	-51	-0.36	2601	0.1296	+ 18.36
369	9.00	+72	+1.54	5184	2.3716	+110.88
330	8.50	+33	+1.04	1089	1.0816	+ 34.32
258	6.00	-39	-1.46	1521	2.1316	+ 56.94
315	7.00	+18	-0.46	324	0.2116	- 8.28
246	6.50	-51	-0.96	2601	0.9216	+ 48.96
330	9.00	+33	+1.54	1089	2.3716	+ 50.82
298	7.50	+ 1	+0.04	1	0.0016	+ 0.04
Σ 2673	67.15			14666	10.0489	+326.60

$$\bar{X} = \frac{\Sigma X}{n} = \frac{2673}{9} = 297.0 \text{ ml};$$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{67.15}{9} = 7.46 \text{ L};$$

$$s_X = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} = \sqrt{\frac{14666}{9-1}} = 42.82 \text{ ml ;}$$

$$s_Y = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n-1}} = \sqrt{\frac{10.0489}{9-1}} = 1.12 \text{ L.}$$

(b) The products of $(X - \bar{X})$ and $(Y - \bar{Y})$ of all individuals are totalled to give $\sum (X - \bar{X})(Y - \bar{Y})$, amounting to 326.60. (Table 8.1.)

(c) r is then computed as follows :

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{(n-1)s_X s_Y} = \frac{326.60}{(9-1) \times 42.82 \times 1.12} = + 0.85.$$

2. Alternative method :

$$\text{Cov}(X, Y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{(n-1)} = \frac{326.60}{9-1} = 40.83 ; \quad r = \frac{\text{Cov}(X, Y)}{s_X s_Y} = \frac{40.83}{42.82 \times 1.12} = + 0.85.$$

Student's t is computed from the product-moment r worked out by any of the above two methods.

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(0.85)^2}{9-2}} = 0.199 ; \quad t = \frac{r}{s_r} = \frac{0.85}{0.199} = 4.271 ; \quad df = n - 2 = 7.$$

Two-tail critical t scores ($df = 7$) are quoted from Table B of Appendix.

$$t_{.05(7)} = 2.365 ; \quad t_{.02(7)} = 2.998 ; \quad t_{.01(7)} = 3.499 ; \quad t_{.001(7)} = 5.405.$$

As the computed t is higher than the critical $t_{.01}$, but lower than the critical $t_{.001}$, the probability P of getting the computed r by chances of random sampling is less than 0.01, but higher than 0.001 ($0.01 > P > 0.001$). So, the variables have a significant correlation below the 0.01 level ($P < 0.01$).

Example 8.2.2.

Compute Pearson's r using either (i) the covariance or (ii) the raw scores to find whether or not there is a significant correlation between height (cm) and weight (kg) in the following data from 9 college students.

Student	:	1	2	3	4	5	6	7	8	9
Height (X)	:	165	182	170	162	160	165	170	170	165
Weight (Y)	:	58.5	60.0	52.0	48.5	49.5	59.0	49.0	56.0	58.0

Solution :

1. First method using covariance :

(a) The X and Y scores are totalled separately to give $\sum X$ and $\sum Y$ (Table 8.2).

(b) Each score is squared and the squared scores of each variable are totalled to give $\sum X^2$ and $\sum Y^2$ respectively (Table 8.2).

(c) The paired X and Y scores of each individual are multiplied with each other and all these products are totalled to give $\sum XY$ (Table 8.2).

(d) The SDs of X and Y scores as well as $Cov(X, Y)$ are computed.

$$s_X = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2} = \sqrt{\frac{253343}{9} - \left(\frac{1509}{9}\right)^2} = 6.09 ;$$

$$s_Y = \sqrt{\frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n}\right)^2} = \sqrt{\frac{26910.75}{9} - \left(\frac{490.5}{9}\right)^2} = 4.45,$$

$$Cov(X, Y) = \frac{\sum XY}{n} - \frac{\sum X \sum Y}{n^2} = \frac{82344.5}{9} - \frac{1509 \times 490.5}{(9)^2} = 11.56.$$

(e) Pearson's r is computed using $Cov(X, Y)$, s_X and s_Y .

$$r = \frac{Cov(X, Y)}{s_X s_Y} = \frac{11.56}{6.09 \times 4.45} = + 0.43.$$

2. Alternative method using raw scores directly :

The computations of $\sum X$, $\sum Y$, $\sum X^2$, $\sum Y^2$ and $\sum XY$ are done as in steps (a) to (c) given in the first method (see above and Table 8.2).

Table 8.2. Table for computing r from covariance and also from raw scores.

	X	Y	X^2	Y^2	XY
	165	58.5	27225	3422.25	9652.5
	182	60.0	33124	3600.00	10920.0
	170	52.0	28900	2704.00	8840.0
	162	48.5	26244	2352.25	7857.0
	160	49.5	25600	2450.25	7920.0
	165	59.0	27225	3481.00	9735.0
	170	49.0	28900	2401.00	8330.0
	170	56.0	28900	3136.00	9520.0
	165	58.0	27225	3364.00	9570.0
Total	1509	490.5	253343	26910.75	82344.5

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{9 \times 82344.5 - 1509 \times 490.5}{\sqrt{[9 \times 253343 - (1509)^2][9 \times 26910.75 - (490.5)^2]}} = + 0.43.$$

Significance of computed r :

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(0.43)^2}{9-2}} = 0.341 ; \quad t = \frac{r}{s_r} = \frac{0.43}{0.341} = 1.261 ; \quad df = n - 2 = 7.$$

On referring to Table B of Appendix, it is seen that even the two-tail critical $t_{.05(7)}$ of 2.365 exceeds the computed t . So, there is *no significant correlation* between the variables ($P > 0.05$).

Example 8.2.3.

Find whether or not there is a significant correlation between vocabulary test scores and typewriting test scores, using the following data.

Student	:	1	2	3	4	5	6	7	8	9	10
Vocabulary score	:	8	22	35	19	23	13	2	14	11	25
Typewriting score	:	29	48	55	49	53	41	22	38	35	48

Solution :

1. First method using the sums of squares :

(a) The means of vocabulary (X) and typewriting (Y) scores, the squared deviations, viz., $(X - \bar{X})^2$ and $(Y - \bar{Y})^2$, of these scores from the respective means, and the products $(X - \bar{X})(Y - \bar{Y})$ of these deviations of scores are worked out (Table 8.3).

$$\bar{X} = \frac{\sum X}{n} = \frac{172}{10} = 17.2 ; \quad \bar{Y} = \frac{\sum Y}{n} = \frac{418}{10} = 41.8.$$

Table 8.3. Table for computing r using the sums of squares.

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
8	29	- 9.2	- 12.8	84.64	163.84	+ 117.76
22	48	+ 4.8	+ 6.2	23.04	38.44	+ 29.76
35	55	+ 17.8	+ 13.2	316.84	174.24	+ 234.96
19	49	+ 1.8	+ 7.2	3.24	51.84	+ 12.96
23	53	+ 5.8	+ 11.2	33.64	125.44	+ 64.96
13	41	- 4.2	- 0.8	17.64	0.64	+ 3.36
2	22	- 15.2	- 19.8	231.04	392.04	+ 300.96
14	38	- 3.2	- 3.8	10.24	14.44	+ 12.16
11	35	- 6.2	- 6.8	38.44	46.24	+ 42.16
25	48	+ 7.8	+ 6.2	60.84	38.44	+ 48.36
$\Sigma 172$	418			819.60	1045.60	+ 867.40

(b) r is computed using the sum of the products of deviations, and the sums of squares (sums of the squared deviations) of variables.

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{867.40}{\sqrt{819.60 \times 1045.60}} = + 0.94.$$

2. Alternative method using covariance :

Alternatively $Cov(X, Y)$, s_X and s_Y may be computed and used in computing r .

$$s_X = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} = \sqrt{\frac{819.60}{10-1}} = 9.54 ; \quad s_Y = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n-1}} = \sqrt{\frac{1045.60}{10-1}} = 10.78 ;$$

$$\text{Cov}(X,Y) = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{(n-1)} = \frac{867.40}{10-1} = 96.38 ; \quad r = \frac{\text{Cov}(X,Y)}{s_X s_Y} = \frac{96.38}{9.54 \times 10.78} = +0.94.$$

Student's t is computed from the product-moment r worked out by either of the two methods, to test the H_0 of no correlation.

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(0.94)^2}{10-2}} = 0.121 ; \quad t = \frac{r}{s_r} = \frac{0.94}{0.121} = 7.769 ; \quad df = n-2 = 10-2 = 8.$$

On consulting Table B of *Appendix*, the computed t is found to be higher than even the two-tail critical $t_{.001(8)}$ which amounts to 5.041. So, the H_0 is rejected — the variables have a *significant correlation* ($P < 0.001$).

Example 8.2.4.

Find whether or not there is a significant correlation between the following gill weights (mg) and body weights (g) of a sample of 10 crabs.

Individuals	:	1	2	3	4	5	6	7	8	9	10
Gill weights	:	70	90	120	160	200	220	220	232	300	310
Body weights	:	3.90	4.82	13.60	14.40	14.82	15.20	15.40	16.11	14.92	16.72

Solution :

1. *First method using raw scores :*

(a) The gill weight (X mg) and the body weight (Y g) scores are totalled to give $\sum X$ and $\sum Y$, respectively (Table 8.4).

(b) Each score is squared and the squared scores of each variable are totalled to give $\sum X^2$ and $\sum Y^2$, respectively (Table 8.4).

(c) The paired X and Y scores of each individual are multiplied with each other to give XY and all these products are totalled give $\sum XY$ (Table 8.4).

Table 8.4. Table for computing r using raw scores.

Individuals	X	Y	X^2	Y^2	XY
1	70	3.90	4900	15.2100	273.00
2	90	4.82	8100	29.2324	433.80
3	120	13.60	14400	184.9600	1632.00
4	160	14.40	25600	207.3600	2304.00
5	200	14.82	40000	219.6324	2964.00
6	220	15.20	48400	231.0400	3344.00
7	220	15.40	48400	237.1600	3388.00
8	232	16.11	53824	259.5321	3737.52
9	300	14.92	90000	222.6064	4476.00
10	310	16.72	96100	279.5584	5183.20
Σ	1922	129.89	429724	1886.2917	27735.52

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{10 \times 27735.52 - 1922 \times 129.89}{\sqrt{[10 \times 429724 - (1922)^2][10 \times 1886.2917 - (129.89)^2]}} = + 0.80.$$

2. Alternative method using sums of squares :

(a) The means of gill weight (X) and body weight (Y) scores, the squared deviations of these scores from the respective means, and the product of these deviations of scores for each individual are worked out (Table 8.5).

$$\bar{X} = \frac{\sum X}{n} = \frac{1922}{10} = 192.2 \text{ mg}; \quad \bar{Y} = \frac{\sum Y}{n} = \frac{129.89}{10} = 12.99 \text{ g.}$$

(b) The squared deviations of scores from their respective means are totalled to give the respective sums of squares, viz., $\sum(X - \bar{X})^2$ and $\sum(Y - \bar{Y})^2$. From Table 8.5 :

$$\sum(X - \bar{X})^2 = 60315.60 \text{ mg}^2; \quad \sum(Y - \bar{Y})^2 = 193.1505 \text{ g}^2.$$

(c) The products of the deviations of scores of the two variables from their respective means are totalled for all the individuals to give the sum of products, viz., $\sum(X - \bar{X})(Y - \bar{Y})$. From Table 8.5 :

$$\sum(X - \bar{X})(Y - \bar{Y}) = 2770.662.$$

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}} = \frac{2770.662}{\sqrt{60315.60 \times 193.1505}} = + 0.81.$$

Table 8.5. Table for computing r using sums of squares.

X	Y	$X - \bar{X}$	$(X - \bar{X})^2$	$Y - \bar{Y}$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
70	3.90	- 122.2	14932.84	- 9.09	82.6281	+ 1110.798
90	4.82	- 102.2	10444.84	- 8.17	66.7489	+ 834.974
120	13.60	- 72.2	5212.84	+ 0.61	0.3721	- 44.042
160	14.40	- 32.2	1036.84	+ 1.41	1.9881	- 45.402
200	14.82	+ 7.8	60.84	+ 1.83	3.3489	+ 14.274
220	15.20	+ 27.8	772.84	+ 2.21	4.8841	+ 61.438
220	15.40	+ 27.8	772.84	+ 2.41	5.8081	+ 66.998
232	16.11	+ 39.8	1584.04	+ 3.12	9.7344	+ 124.176
300	14.92	+ 107.8	11620.84	+ 1.93	3.7249	+ 208.054
310	16.72	+ 117.8	13876.84	+ 3.73	13.9129	+ 439.394
Σ 1922	129.89		60315.60		193.1505	+ 2770.662

To test the H_0 that there is no significant correlation, r is worked out by either of the two methods and converted to Student's t .

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(0.81)^2}{10-2}} = 0.207 ; \quad t = \frac{r}{s_r} = \frac{0.81}{0.207} = 3.913 ; \quad df = n - 2 = 10 - 2 = 8.$$

Using Table B of *Appendix*, the computed t is found to be lower than the critical $t_{.001(8)}$ of 5.041, but to be higher than the critical $t_{.01(8)}$ of 3.355. So, the probability P of correctness of the H_0 is higher than 0.001, but lower than 0.01. Because $P < 0.01$, the H_0 is rejected and it is inferred that there is a *significant correlation* between gill weights and body weights.

Example 8.2.5.

(1) Find whether or not there is a significant correlation between the IQ values of the following 10 students and the marks obtained by them in Mathematics in a school examination.

Individuals	:	1	2	3	4	5	6	7	8	9	10
IQ values	:	95	102	112	96	134	125	100	108	110	88
Maths marks	:	20	43	45	30	65	60	46	50	57	24

(2) Find whether or not the correlation coefficient obtained in this sample differs significantly from a parametric correlation coefficient of + 0.60.

Solution :

(1) *Test of the H_0 of no correlation :*

In this case, the H_0 contends that there is no significant correlation between the two variables, which means in effect that the population correlation is proposed by the H_0 to be zero : ($H_0 : \rho = 0$; $H_a : \rho \neq 0$). So, this H_0 can be tested by converting the computed r to Student's t score.

(a) The means of IQ values (X) and Maths marks (Y) are worked out and used to compute the deviations of the scores of each variable from the respective means (Table 8.6). These deviations are next squared and these squared deviations for each variable are totalled to give the respective sums of squares, viz., $\Sigma(X - \bar{X})^2$ and $\Sigma(Y - \bar{Y})^2$. Using Table 8.6,

$$\bar{X} = \frac{\Sigma X}{n} = \frac{1070}{10} = 107.0 ; \quad \bar{Y} = \frac{\Sigma Y}{n} = \frac{440}{10} = 44.0.$$

$$\Sigma(X - \bar{X})^2 = 1788 ; \quad \Sigma(Y - \bar{Y})^2 = 2080.$$

(b) For each individual, the product of $(X - \bar{X})$ and $(Y - \bar{Y})$ is worked out. All such products are totalled to give the sum of products, viz., $\Sigma(X - \bar{X})(Y - \bar{Y})$. From Table 8.6,

$$\Sigma(X - \bar{X})(Y - \bar{Y}) = 1718.$$

$$\therefore r = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2 \Sigma(Y - \bar{Y})^2}} = \frac{1718}{\sqrt{1788 \times 2080}} = + 0.89.$$

Table 8.6. Table for computing r using sums of squares.

IQ values (X)	Maths marks (Y)	$X - \bar{X}$	$(X - \bar{X})^2$	$Y - \bar{Y}$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
	20	-12	144	-24	576	+ 288
95	43	- 5	25	- 1	1	+ 5
102	45	+ 5	25	+ 1	1	+ 5
112	30	-11	121	-14	196	+ 154
96	65	+27	729	+21	441	+ 567
134	60	+18	324	+16	256	+ 288
125	46	- 7	49	+ 2	4	- 14
100	50	+ 1	1	+ 6	36	+ 6
108	57	+ 3	9	+13	169	+ 39
110	24	-19	361	-20	400	+ 380
88						
Σ 1070	440		1788		2080	+1718

The computed r is converted to Student's t , using the SE (s_r) of r .

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(0.89)^2}{10-2}} = 0.161; \quad t = \frac{r}{s_r} = \frac{0.89}{0.161} = 5.528; \quad df = n - 2 = 10 - 2 = 8.$$

Using Table B of Appendix, the computed t is found to be higher than even the critical $t_{.001(8)}$ of 5.041. So, the probability (P) of the H_0 being correct is too low — lower than 0.001. Hence, the H_0 is rejected and it is inferred that there is a significant correlation between IQ and Maths marks ($P < 0.001$).

2. Test of the H_0 proposing ρ other than zero :

In this case, the H_0 contends that the computed r is not significantly different from the proposed parametric correlation coefficient (ρ) amounting to 0.60.

$$H_0 : \rho = + 0.60; \quad H_a : \rho \neq 0.60.$$

In such a case, Fisher's z transformation has to be used in working out the probability (P) of the H_0 being correct (pages 141-142).

(a) The computed r and the proposed ρ are both logarithmically transformed respectively into z_r and ζ .

$$z_r = 1.1513 \log \left[\frac{1+r}{1-r} \right] = 1.1513 \log \left[\frac{1+0.89}{1-0.89} \right] = 1.422.$$

$$\zeta = 1.1513 \log \left[\frac{1+\rho}{1-\rho} \right] = 1.1513 \log \left[\frac{1+0.60}{1-0.60} \right] = 0.693.$$

(b) The difference between z_r and ζ is converted to the standard z score, using the SE of the difference (s_{z_r}). The probability (P) of the computed z occurring by chance is then worked out using the unit normal curve table (Table A of Appendix).

$$s_{z_r} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{10-3}} = 0.3780; \quad z = \frac{z_r - \zeta}{s_{z_r}} = \frac{1.422 - 0.693}{0.3780} = 1.93.$$

$$P = 2 [0.5000 - (\text{area of unit normal curve from its } \mu \text{ to the computed } z)] \\ = 2 [0.5000 - 0.4732] = 0.054.$$

Because the P of correctness of the H_0 exceeds 0.05, P is considered too high and the H_0 is retained. It is, therefore, inferred that the computed r does not differ significantly from the proposed ρ of + 0.60 ($P > 0.05$).

Computation from grouped data

A code method is used in computing r from a bivariate frequency distribution of a large sample.

Bivariate frequency distribution :

It is the frequency distribution of the paired scores of two variables for the individuals of a sample. It may be either arranged in a two-way table as follows, or represented graphically as a scattergram (pages 33-34), or drawn as a three-dimensional bell-shaped/elliptical mound-like distribution (Fig. 8.3.).

(a) Each variable is divided into a number of classes as in the case of univariate frequency distributions (pages 15-16).

(b) The classes of the variable Y are arranged in a descending order from top to bottom in the rows of a table while those of the variable X are arranged in an ascending order from left to right in the columns (Table 8.7). A two-way table results with k and l numbers of classes for X and Y respectively, and $k \times l$ number of cells. Thus, the bottom left corner cell accommodates the frequency of individuals with scores of both variables falling in their respective lowest classes while the top right corner cell houses the frequency of those with scores in the highest classes of both variables.

(c) According to his scores in the two variables, each individual is entered as a tally in the cell corresponding to the relevant combination of class intervals of the respective variables.

(d) The total number of tallies in each cell is entered at its centre as the cell frequency.

(e) The sum of all cell frequencies of each

row is entered as the *marginal frequency of the row* (f_Y) in the extreme right column of the table. The f_Y values in this column constitute the frequency distribution of Y scores (*marginal distribution of Y*).

(f) The sum of all cell frequencies of each column is entered as the *marginal frequency of the column* (f_X) in the bottom row of the table. The f_X values in this row constitute the frequency distribution of X scores (*marginal distribution of X*).

$$\Sigma f_Y = \Sigma f_X = n.$$

Code method for computing r :

(a) The bivariate frequency distribution is used in a correlation table with its cell frequencies at the centres of the respective cells (Table 8.8). The marginal frequency (f_Y) of each row and the marginal frequency (f_X) of each column are entered in respectively the marginal column and row for those values. Additional marginal columns and rows are added to the right of the f_Y column and below the f_X row, respectively.

(b) The midpoint of a centrally located class interval of each variable is chosen as its *assumed mean* (A) and given a code number of 0 (Example 8.2.6).

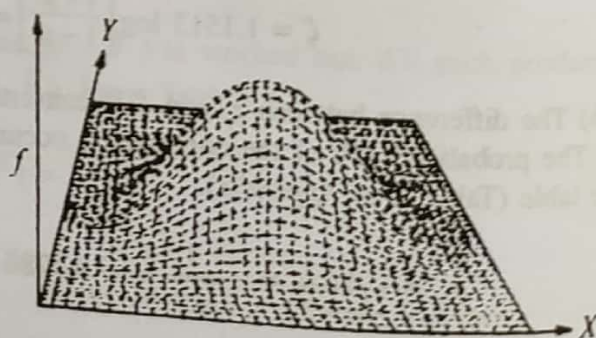


Fig. 8.3. Bivariate normal frequency distribution.

Table 8.7. Tabulation of a bivariate frequency distribution.

		X variable : Age in years							f_Y
		19.5-27.5	27.5-35.5	35.5-43.5	43.5-51.5	51.5-59.5	59.5-67.5	67.5-75.5	
Y variable : Cardiac index in litres	3.42-3.59	III 5							5
	3.25-3.42	I 1	III 5	III 3	II 2				11
	3.08-3.25		III 3	II 2	III 3	II 2			10
	2.91-3.08		III 3	II 2	III 5	III 4	I 1		15
	2.74-2.91			III 3	III 5	III 4	III 5	I 1	18
	2.57-2.74				II 2	III 4	III 5	I 1	12
	2.40-2.57					II 2	II 2	III 5	9
f_X		6	11	10	17	16	13	7	80 (n)

(c) Code numbers (x' and y') are assigned to the midpoints of other intervals of X and Y variables, respectively, according to the orders and the directions (signs) of the deviations of those midpoints from the respective A values. For example, the class interval of X just below that of its A ($x' = 0$) bears the x' number of -1; the interval of Y just higher than that of its A ($y' = 0$) is given the y' number of +1. These code numbers are entered in the marginal row for x' and the marginal column for y' respectively.

(d) The f_X of each column is multiplied by the x' of that column to give fx' which is again multiplied by x' to yield fx'^2 of that column. The sums of fx' and fx'^2 values of all the columns give respectively $\Sigma fx'$ and $\Sigma fx'^2$.

(e) The f_Y and y' values of each row are similarly used in computing fy' , fy'^2 , $\Sigma fy'$ and $\Sigma fy'^2$ values.

(f) The means (\bar{x}' and \bar{y}') and the SDs (s'_x and s'_y) of the coded values of X and Y are then computed.

$$\bar{x}' = \frac{\Sigma fx'}{n}; \quad \bar{y}' = \frac{\Sigma fy'}{n};$$

$$s'_x = \frac{1}{n} \sqrt{n \Sigma fx'^2 - (\Sigma fx')^2};$$

$$s'_y = \frac{1}{n} \sqrt{n \Sigma fy'^2 - (\Sigma fy')^2}.$$

(g) The product of x' and y' of each cell is entered at its top left corner. Each such $x'y'$ value is multiplied by the frequency of that cell, this product is entered at its bottom right corner, and such products of all the cells in each row (or column) are added to give $\Sigma x'y'_{\text{row}}$ (or $\Sigma x'y'_{\text{column}}$). All entries of $\Sigma x'y'_{\text{row}}$ (or $\Sigma x'y'_{\text{column}}$) are added to give $\Sigma x'y'$.

$$r = \frac{\Sigma x'y' - n\bar{x}'\bar{y}'}{ns'_x s'_y}.$$

Example 8.2.6.

Use the bivariate frequency distribution of Table 8.7 to compute r between age (yrs) and cardiac index (litres). Interpret the result in terms of the H_0 proposing a population correlation coefficient of -0.85 .

Solution :

The bivariate frequency distribution of Table 8.7 is rearranged in Table 8.8 for computing r by the code method.

(a) The midpoint of the interval $43.5 - 51.5$ is chosen as the assumed mean A of X (age) and given the code number (x') of 0. The midpoint of $2.91 - 3.08$ is chosen as A of Y (cardiac index) and given the code number (y') of 0.

(b) The midpoints of other intervals of X and Y are then given respectively x' and y' code numbers according to their deviations from the respective assumed means (A).

(c) Using the f_x and x' values of each column of Table 8.8, fx' and fx'^2 are computed for that column. The f_y and y' of each row are similarly used in computing its fy' and fy'^2 .

(d) The means (\bar{x}' and \bar{y}') and the SDs (s_x' and s_y') are computed for the coded values of X and Y .

$$\bar{x}' = \frac{\sum fx'}{n} = \frac{13}{80} = 0.16 ; \quad \bar{y}' = \frac{\sum fy'}{n} = \frac{-22}{80} = -0.28 ;$$

$$s_x' = \frac{1}{n} \sqrt{n \sum fx'^2 - (\sum fx')^2} = \frac{1}{80} \sqrt{80 \times 239 - (13)^2} = 1.721 ;$$

$$s_y' = \frac{1}{n} \sqrt{n \sum fy'^2 - (\sum fy')^2} = \frac{1}{80} \sqrt{80 \times 246 - (-22)^2} = 1.732.$$

(e) The product of x' and y' of each cell is entered at its top left corner ; e.g., for the bottom right corner cell of the table, $x'y' = 3 \times (-3) = -9$. Each $x'y'$ is multiplied by its cell frequency and that product is entered at the bottom right corner of the cell ; e.g., for the above-mentioned cell, this value comes to : $fx'y' = 5 \times (-9) = -45$. The last-mentioned products are totalled for all cells in each row (or column) to give $\sum x'y'_{\text{row}}$ (or $\sum x'y'_{\text{column}}$). All $\sum x'y'_{\text{row}}$ (or $\sum x'y'_{\text{column}}$) values are totalled to give $\sum x'y'$ which amounts to -194 .

(f) r is then computed using the values of Table 8.8.

$$r = \frac{\sum x'y' - n\bar{x}'\bar{y}'}{ns_x's_y'} = \frac{-194 - 80 \times 0.16 \times (-0.28)}{80 \times 1.721 \times 1.732} = -0.799.$$

(g) To test the H_0 that the computed r is not significantly different from the proposed ρ of -0.85 , both r and ρ are subjected to Fisher's z transformation into z_r and ζ respectively. The computed z_r is then converted to the standard z score.

$$z_r = 1.1513 \log \left[\frac{1+r}{1-r} \right] = 1.1513 \log \left[\frac{1-0.799}{1+0.799} \right] = -1.096.$$

$$\zeta = 1.1513 \log \left[\frac{1+\rho}{1-\rho} \right] = 1.1513 \log \left[\frac{1-0.85}{1+0.85} \right] = -1.256.$$

$$s_{z_r} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{80-3}} = 0.1140 ; \quad z = \frac{z_r - \zeta}{s_{z_r}} = \frac{-1.096 + 1.256}{0.1140} = 1.40.$$

Table A of unit normal curve areas is then used to find the two-tail probability P for the H_0 being correct.

$$P = 2 [0.5000 - (\text{area of unit normal curve from its } \mu \text{ to the computed } z \text{ of } 1.40)] \\ = 2 [0.5000 - 0.4192] = 0.16.$$

As $P > 0.05$, it is considered too high and the H_0 cannot be rejected. So, the computed r does not differ significantly from the given ρ of -0.85 ($P > 0.05$).

Significance of difference between r scores

Fisher's z transformation (pages 141-142) is used to find whether or not there is a significant difference between the product moment correlation coefficients (r_1 and r_2) of two given variables (X_1 and X_2) in two different samples, each consisting of more than 50 cases ($n_1 > 50$, $n_2 > 50$). The H_0 contends that there is no significant difference between the two r values; in other words, H_0 proposes that the parametric correlation coefficients, ρ_1 and ρ_2 , of which r_1 and r_2 may serve as the respective estimates, are identical.

$$H_0 : \rho_1 = \rho_2 ; \quad H_a : \rho_1 \neq \rho_2.$$

To work out the probability (P) of this H_0 being correct, each of the sample r values is logarithmically converted to the corresponding z_r and the difference ($z_{r_1} - z_{r_2}$) between the two z_r values is converted to the standard z score using the SE (s_{z_r}) of the difference between them.

$$z_{r_1} = 1.1513 \log \left[\frac{1+r_1}{1-r_1} \right] ;$$

$$z_{r_2} = 1.1513 \log \left[\frac{1+r_2}{1-r_2} \right] ;$$

$$s_{z_r} = \sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}} ; \quad z = \frac{z_{r_1} - z_{r_2}}{s_{z_r}}.$$

The probability (P) of correctness of the H_0 is obtained using the fractional area of the unit normal curve (Table A of Appendix).

$$P = 2 [0.5000 - (\text{area of unit normal curve from its } \mu \text{ to the computed } z)].$$

If the P thus worked out is found not to exceed 0.05 or any other chosen significance level, P is considered too low; the H_0 is then rejected and the two computed r values are considered to differ significantly ($P \leq \alpha$). But if the computed P exceeds 0.05 or any other chosen α , P is considered too high, the H_0 cannot be rejected, and the two r values are considered not to differ significantly ($P > \alpha$).

Example 8.2.7.

The product-moment r between trunk lengths and wing lengths amounted respectively to $+0.63$ and $+0.52$ in two samples of respectively 57 and 60 cockroaches sampled from two different habitats. Find whether or not there is a significant difference between the two sample r values ($\alpha = 0.01$).

Solution :

$$\text{Sample 1 : } r_1 = +0.63 ; \quad n_1 = 57.$$

$$\text{Sample 2 : } r_2 = +0.52 ; \quad n_2 = 60.$$

Both r_1 and r_2 are converted to the respective z_r values ; the $SE (s_{z_r})$ of the difference between the z_r coefficients is worked out and used in transforming the difference between the z_r values to the standard z score.

$$z_{r_1} = 1.1513 \log \left[\frac{1+r_1}{1-r_1} \right] = 1.1513 \log \left[\frac{1+0.63}{1-0.63} \right] = 0.741 ;$$

$$z_{r_2} = 1.1513 \log \left[\frac{1+r_2}{1-r_2} \right] = 1.1513 \log \left[\frac{1+0.52}{1-0.52} \right] = 0.576 ;$$

$$s_{z_r} = \sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}} = \sqrt{\frac{1}{57-3} + \frac{1}{60-3}} = 0.1899 ;$$

$$z = \frac{z_{r_1} - z_{r_2}}{s_{z_r}} = \frac{0.741 - 0.576}{0.1899} = 0.87.$$

The probability (P) of the H_0 being correct is worked out using the unit normal curve table (Table A of Appendix).

$$\begin{aligned} P &= 2 [0.5000 - (\text{area of unit normal curve from its } \mu \text{ to the computed } z)] \\ &= 2 [0.5000 - 0.3078] = 0.38. \end{aligned}$$

As the computed P exceeds the chosen α of 0.01, it is considered too high. So, the H_0 is retained, and it is inferred that there is *no significant difference* between the r values of the two samples ($P > 0.01$).

Example 8.2.8.

The product-moment r values between the intelligence test scores and the numerical reasoning test scores amounted respectively to +0.65 and +0.50 in two samples of 83 and 54 students, respectively, from two different educational institutions. Is there a significant difference between the r values of the two samples ?

Solution :

Sample 1 : $r_1 = + 0.65$; $n_1 = 83$.

Sample 2 : $r_2 = + 0.50$; $n_2 = 54$.

Both r_1 and r_2 are converted to z_r and the difference between the computed z_r values is changed into the standard z score, using the $SE (s_{z_r})$ of their differences. The computed standard z score is then used to work out the probability (P) of correctness of the H_0 of no difference.

$$z_{r_1} = 1.1513 \log \left[\frac{1+r_1}{1-r_1} \right] = 1.1513 \log \left[\frac{1+0.65}{1-0.65} \right] = 0.775 ;$$

$$z_{r_2} = 1.1513 \log \left[\frac{1+r_2}{1-r_2} \right] = 1.1513 \log \left[\frac{1+0.50}{1-0.50} \right] = 0.549 ;$$

$$s_{z_r} = \sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}} = \sqrt{\frac{1}{83-3} + \frac{1}{54-3}} = 0.1792 ;$$

$$z = \frac{z_{r_1} - z_{r_2}}{s_{z_r}} = \frac{0.775 - 0.549}{0.1792} = 1.26.$$

$$P = 2 [0.5000 - (\text{area of unit normal curve from its } \mu \text{ to the computed } z)] \\ = 2 [0.5000 - 0.3962] = 0.21.$$

As the probability (P) of the H_0 being correct is higher than 0.05, it is considered too high. So, the H_0 is retained, and it is inferred that there is *no significant difference* between the r values of the two samples ($P > 0.05$).

8.3 PARTIAL CORRELATION

Partial correlation, though a special form of correlation between two given variables, is a method of *multivariate statistics* involving more than two variables in a sample. The product-moment r , though intended to measure the simple correlation between two given variables, results partly from the effects of other variables on both of them. Partial correlation aims at eliminating (partialling out) such effects of the other variables on both the given variables in common so as to get a more precise measure of the correlation directly between the given variables. It is that part of the product-moment r between two given variables, which remains after the elimination of the component of their association, arising from the effects of the other variables on both of them. There are different *orders of partial correlation* according to the number of variables to be eliminated or partialled out; the product-moment r between any two variables, involving no elimination of any other variable, is considered as the *zero-order r* .

Assumptions

For using partial r , it should be reasonable to assume that,

- (a) all the variables involved are *continuous measurement variables*;
- (b) their scores have *unimodal and fairly symmetrical distributions* in the population, without marked skewness;
- (c) the paired scores of each pair of variables in an individual or case are *independent* of such paired scores of all other

individuals or cases in the sample;

(d) there is a *linear association* between the scores of each pair of variables.

First-order partial r

This is the correlation between two variables (X_1 and X_2), partialling out another variable (X_3) correlated with both of them. Thus, of the three correlated variables, the first-order partial r holds one constant to remove its effect on the correlation between the other two. For example, $r_{12.3}$ partials out the variable X_3 to measure the correlation between X_1 and X_2 , free from the effect of X_3 . Here, any X_1 score consists of two independent and uncorrelated components, one correlated with X_3 and the other independent of X_3 ; similarly, each X_2 score has one component correlated with X_3 and another component independent of the latter. The zero-order r (r_{12}) between X_1 and X_2 is partly due to their respective components correlated with X_3 ; $r_{12.3}$ partials out or eliminates these components associated with X_3 and measures the correlation between the residual components of X_1 and X_2 , free from the influence of X_3 .

The partial r of any order is computed using the product-moment r and the lower orders of partial r . For a first-order partial r ,

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

where r_{12} , r_{13} and r_{23} are zero-order product-moment r values between respectively X_1 and X_2 , X_1 and X_3 , and X_2 and X_3 . If, in a sample of humans, r_{12} is the correlation between

cardiac output (X_1) and venous return (X_2), r_{13} is the correlation between X_1 and blood pressure (X_3), and r_{23} is the correlation between X_2 and X_3 , $r_{12.3}$ is the partial r between cardiac output and venous return, eliminating the effects of blood pressure; similarly, if intelligence test scores (X_1), anxiety test scores (X_2) and age (X_3) are correlated with each other in a sample of students, $r_{12.3}$ is the partial r between intelligence and anxiety scores, eliminating the effects of age; again, $r_{12.3}$ is the first-order partial r between oxygen consumption (X_1) and tracheal ventilation (X_2), partialling out the effects of atmospheric sulfur di-oxide concentration (X_3), in a sample of locusts. Partial r values range from -1.00 to $+1.00$.

Other first-order partial r values may also be computed similarly.

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}};$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1-r_{12}^2)(1-r_{13}^2)}}.$$

For example, in one case cited above, $r_{13.2}$ is the partial r between intelligence test scores (X_1) and age (X_3), partialling out anxiety test scores (X_2). Similarly, $r_{23.1}$ is the partial r between the rate of plant stem growth (X_2) and the soil pH (X_3), eliminating the effects of soil moisture (X_1), when the variables are correlated with each other.

The null hypothesis (H_0) proposes that the partial correlation between the given variables amounts to zero in the population, and that the

computed partial r has resulted only due to chances associated with random sampling. To test the probability P of this H_0 being correct, the computed partial r is transformed into the t score and the latter is compared with critical t scores having the same df . For $r_{12.3}$,

$$s_{r_{12.3}} = \sqrt{\frac{(1-r_{12.3}^2)}{n-3}}; \quad t = \frac{r_{12.3}}{s_{r_{12.3}}};$$

$$df = n - 3.$$

The computed partial r is considered significant only if the t score worked out from it either exceeds or equals the critical t score for a chosen level of significance, not higher than 0.05 ($P \leq \alpha$).

Second-order partial r

The second and still higher orders of partial r are seldom used. The *second-order partial r* involves four inter-correlated variables and measures the correlation between two of them, partialling out the other two. For example, $r_{12.34}$ correlates X_1 and X_2 , eliminating the effects of X_3 and X_4 . The significance of the computed $r_{12.34}$ is tested by transforming it to t score and comparing the latter to critical t scores.

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}r_{24.3}}{\sqrt{(1-r_{14.3}^2)(1-r_{24.3}^2)}};$$

$$\text{or, } r_{12.34} = \frac{r_{12.4} - r_{13.4}r_{23.4}}{\sqrt{(1-r_{13.4}^2)(1-r_{23.4}^2)}};$$

$$s_{r_{12.34}} = \frac{1}{\sqrt{n-4}}; \quad t = \frac{r_{12.34}}{s_{r_{12.34}}}; \quad df = n - 4.$$

Example 8.3.1.

In a group of 153 students, the product-moment r values between intelligence test scores (X_1), anxiety test scores (X_2) and chronological age (X_3) were found to be as follows: $r_{12} = +0.46$, $r_{13} = +0.35$, $r_{23} = +0.17$. Compute $r_{12.3}$ and find whether it is significant or not.

Solution :

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = \frac{0.46 - 0.35 \times 0.17}{\sqrt{(1-0.35^2)(1-0.17^2)}} = + 0.434 ;$$

$$t = \frac{r_{12.3}}{\sqrt{(1-r_{12.3}^2)/(n-3)}} = \frac{0.434}{\sqrt{\frac{1-(0.434)^2}{153-3}}} = 5.9000 ;$$

$$df = n - 3 = 153 - 3 = 150 = \infty.$$

For a two-tail test, critical $t_{.01(\infty)} = 2.576$, and critical $t_{.001(\infty)} = 3.291$ (Table B). The computed t is thus higher than the critical t for the 0.001 level. So, the partial r is *significant* ($P < 0.001$).

Example 8.3.2.

In a sample of 63 athletes, the product-moment r values between the stroke volume of heart (X_1), the venous return (X_2) and the vascular peripheral resistance (X_3) were found to be as follows : $r_{12} = + 0.65$, $r_{13} = - 0.12$, $r_{23} = - 0.25$. Compute the partial r between stroke volume and venous return, eliminating the effect of peripheral resistance, and test its significance.

Solution :

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = \frac{0.65 - (-0.12) \times (-0.25)}{\sqrt{[1-(-0.12)^2][1-(-0.25)^2]}} = + 0.645 ;$$

$$t = \frac{r_{12.3}}{\sqrt{(1-r_{12.3}^2)/(n-3)}} = \frac{0.645}{\sqrt{\frac{1-(0.645)^2}{63-3}}} = 6.538 ; \quad df = n - 3 = 63 - 3 = 60.$$

For a two-tail test, critical $t_{.01(60)} = 2.660$, and critical $t_{.001(60)} = 3.460$ (Table B). The computed t is thus higher than the critical t for the 0.001 level of significance. So, the partial r is *significant* ($P < 0.001$).

8.4 MULTIPLE CORRELATION

Multiple correlation comes under *multi-variate statistics* as it involves more than two variables in a sample. It is a measure of the relation between one variable (called the *criterion*) and the weighted sum of two or more other variables (called the *predictors*). The *multiple linear correlation coefficient* (R) is a special form of the product-moment r and measures the magnitude of the linear relationship between a given variable and the

weighted sum of two or more other variables.

Assumptions

For computing and applying R , it should be reasonable to assume that :

(a) all the variables involved, both *criterion* and *predictors*, are *continuous measurement variables* ;

(b) their scores have *unimodal* and *fairly symmetrical distributions* in the population, without marked skewness ;

(c) the paired scores of each pair of variables in an individual or case occur in the sample *at random* and *independent* of such paired scores of all other individuals or cases ;

(d) there is a *linear association* between the scores of each pair of variables.

Multiple correlation with three variables

Such a multiple linear correlation coefficient ($R_{1.23}$) may be computed between a criterion variable (X_1) and a weighted sum of two predictor variables (X_2 and X_3) using the beta coefficients, β_2 and β_3 . β_2 and β_3 are those proportions of the total variance of the criterion X_1 as are associated with the variances of respectively X_2 and X_3 . Where r_{12} , r_{13} and r_{23} are the product-moment r values between X_1 and X_2 , X_1 and X_3 , and X_2 and X_3 respectively,

$$\beta_2 = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2}; \quad \beta_3 = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2};$$

$$R_{1.23} = \sqrt{\beta_2 r_{12} + \beta_3 r_{13}};$$

$$\text{or, } R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}.$$

$R_{2.13}$ and $R_{3.12}$ are also computed similarly.

$$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{13}^2}};$$

$$R_{3.12} = \sqrt{\frac{r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{12}^2}}.$$

If the multiple correlation is perfectly linear, R amounts to 1.00 exactly. If R amounts to 0, linear multiple correlation does not exist, but a nonlinear correlation cannot be straightway negated.

The magnitude of the computed $R_{1.23}$ tends to be higher than the product-moment r_{12} between the criterion (X_1) and a predictor (X_2)

in the following cases :

(i) if the product-moment r_{23} between the two predictors (X_2 and X_3) is low ;

(ii) if the criterion has a high r_{12} or r_{13} with either predictor ;

(iii) if r_{12} is substantially decreased by the negative effect of the other predictor X_3 , called the *suppression variable*, because of a high r_{23} and a negative, poor or zero r_{13} . This happens when the suppression variable X_3 has a variance common with the other predictor X_2 , but no such common variance with the criterion X_1 ; there may still exist a positive r_{12} because of a variance common to X_1 and X_2 , but it is partly suppressed by the negative weight of the other variance, common to X_2 and X_3 and absent in X_1 . $R_{1.23}$ attains a value higher than r_{12} by minimizing this negative weight of the suppression variable.

Coefficients of multiple determination and non-determination :

The *coefficient of multiple determination* (R^2) is a measure of that proportion of variance of the criterion which comes from the combined contribution of the predictors. In other words, it is a measure of that proportion of the variation of the criterion, which is determined by the variations of the predictors. R^2 finds application in the analysis of multiple regression and in assessing the relative importance of multiple correlations of different magnitudes. For a multiple correlation between a criterion (X_1) and two predictors (X_2 and X_3),

$$R_{1.23}^2 = \beta_2 r_{12} + \beta_3 r_{13}.$$

That remaining proportion of the variance of the criterion as is independent of the combined contribution of the predictors, is given by the *coefficient of multiple non-determination* (K^2) which equals $1 - R^2$. Thus, for a multiple correlation between the criterion (X_1) and two predictors (X_2 and X_3),

$$K_{1.23}^2 = 1 - R_{1.23}^2.$$

Interpretation :

To test the H_0 that the computed R is not significantly different from 0, R is converted to t which is compared with critical t scores (Table B of Appendix) for interpretation.

$$s_R = \frac{1}{\sqrt{n-3}} ; \quad t = \frac{R}{s_R} ; \quad df = n - 3.$$

Alternatively, the computed R is converted to the variance ratio (F) which is compared with critical F ratios (Table H of Appendix) for the combination of specified df_1 and df_2 (viz., degrees of freedom of respectively *greater and lesser mean squares* in the table).

$$F = \frac{R^2(n-g-1)}{g(1-R^2)} ;$$

$$df_1 = g ; \quad df_2 = n - g - 1 ;$$

where g is the number of predictors. For different multiple correlations with three variables, $df_1 = g = 2$; $df_2 = n - g - 1 = n - 3$;

$$F = \frac{R_{1,23}^2(n-3)}{2(1-R_{1,23}^2)} ; \quad F = \frac{R_{2,13}^2(n-3)}{2(1-R_{2,13}^2)} ;$$

Example 8.4.1

In a sample of 63 persons, the r values between the cardiac stroke volume (X_1), the venous return (X_2) and the vascular peripheral resistance (X_3) were found to be as follows : $r_{12} = +0.65$, $r_{13} = -0.12$, $r_{23} = -0.25$. Compute the multiple correlation coefficient between stroke volume and the combination of venous return and peripheral resistance.

Solution :

(a) Beta coefficients are computed from the r values between the variables, and used in computing the multiple correlation coefficient $R_{1,23}$.

$$\beta_2 = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} = \frac{0.65 - (-0.12) \times (-0.25)}{1 - (-0.25)^2} = 0.661.$$

$$\beta_3 = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} = \frac{-0.12 - 0.65 \times (-0.25)}{1 - (-0.25)^2} = 0.045.$$

$$R_{1,23} = \sqrt{\beta_2 r_{12} + \beta_3 r_{13}} = \sqrt{0.661 \times 0.65 + 0.045 \times (-0.12)} = 0.651.$$

(b) To test the H_0 of no correlation, $R_{1,23}$ is converted to t score which is compared with critical t scores.

$$F = \frac{R_{3,12}^2(n-3)}{2(1-R_{3,12}^2)}.$$

The computed R is significantly different from 0 at or below that level of significance which has a critical F respectively equal to or lower than the computed F ($P \leq \alpha$).

Multiple correlation with more than three variables

Multiple correlation with one criterion and more than two predictors can also be computed using the beta coefficients. Where m is the total number of variables and g is the number of predictors,

$$R_{1,23\dots m} = \sqrt{\beta_2 r_{12} + \beta_3 r_{13} + \dots + \beta_m r_{1m}} ;$$

$$F = \frac{R_{1,23\dots m}^2(n-g-1)}{g(1-R_{1,23\dots m}^2)} ;$$

$$df_1 = g ; \quad df_2 = n - g - 1.$$

$$s_R = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{63-3}} = 0.129.$$

$$t = \frac{R_{1.23}}{s_R} = \frac{0.651}{0.129} = 5.047 ; \quad df = n - 3 = 63 - 3 = 60.$$

For a two-tail test, critical $t_{.001(60)} = 3.460$ (Table B). As the computed t is higher than even this critical t score, the computed $R_{1.23}$ is significantly different from zero ($P < 0.001$).

Alternatively, $R_{1.23}$ is converted to F ratio which is compared with critical F ratios.

$$F = \frac{R_{1.23}^2(n-3)}{2(1-R_{1.23}^2)} = \frac{(0.651)^2(63-3)}{2[1-(0.651)^2]} = 22.065.$$

$$df_1 = g = 2 ; \quad df_2 = n - g - 1 = 63 - 2 - 1 = 60.$$

From Table H of Appendix, critical $F_{.05(2,60)} = 3.15$; critical $F_{.01(2,60)} = 4.98$. As the computed F is much higher than the critical F for 0.01 level of significance, $R_{1.23}$ is significantly different from zero ($P \ll 0.01$).

Example 8.4.2.

In a sample of 153 students, the product-moment r values between intelligence test scores (X_1), anxiety test scores (X_2) and chronological age (X_3) were found to be as follows : $r_{12} = + 0.46$, $r_{13} = + 0.35$, $r_{23} = + 0.17$. Compute and interpret the multiple correlation coefficient between intelligence test scores and the combination of anxiety test scores and age.

Solution :

$$R_{1.23} = \sqrt{\frac{r_{12}^2 - 2r_{12}r_{13}r_{23} + r_{13}^2}{1 - r_{23}^2}} = \sqrt{\frac{(0.46)^2 - 2 \times 0.46 \times 0.35 \times 0.17 + (0.35)^2}{1 - (0.17)^2}} = 0.536.$$

$$F = \frac{R_{1.23}^2(n-3)}{2(1-R_{1.23}^2)} = \frac{(0.536)^2(153-3)}{2[1-(0.536)^2]} = 30.233 ; \quad df_1 = g = 2 ; \quad df_2 = n - g - 1 = 153 - 2 - 1 = 150.$$

From Table H of Appendix, critical $F_{.05(2,150)} = 3.06$; critical $F_{.01(2,150)} = 4.75$. Since the computed F far exceeds the critical F for 0.01 level of significance, $R_{1.23}$ is significantly different from zero ($P \ll 0.01$).

8.5 SPEARMAN'S RANK CORRELATION

Spearman's rank-difference correlation coefficient (ρ or r_ρ) is a *nonparametric* counterpart of the product-moment r for *simple linear correlation*. It explores the magnitude and direction of the linear relation between two variables in a sample when their values are expressed in ranks. The computed r_ρ values

range from -1.00 and $+1.00$, total absence of correlation being indicated by 0.00 . In contrast to the product-moment r , it has fewer and simpler assumptions, and is computed more easily, but is less powerful.

Assumptions

For using Spearman's ρ , it should be reasonable to assume that :

(a) there is a *linear relationship* between the variables to be correlated ;

(b) the magnitude of each variable can be expressed in ranks — r_ρ is thus applicable to ordinal variables and to continuous ratio and interval variables whose scores can be changed into ranks, but not to nominal variables which cannot be ranked ;

(c) the paired ranks (scores) of the variables of each individual or case occur in the sample at *random* and *independent* of all other paired ranks (scores).

As the assumptions for normal distributions and for continuous nature of the variables are not required, it can be applied to both continuous and discontinuous variables, distributed normally or non-normally, as also to small samples.

Computation

(a) Ranks are first assigned to the scores of each variable separately, in either ascending or descending order for both variables. If two or more scores of a variable are identical, each of such tied scores is given an *average rank* which is the arithmetic mean of the ranks that those scores of the tied set would have got if they were successive scores instead of being identical. The score next to a tied set is given the same rank as it would have got if the tied scores of the preceding set would have held separate successive ranks.

However, if the individuals of the sample were already ranked with respect to each of the two variables in similar orders, those ranks can be used instead of fresh ranking.

(b) The difference (D) between the ranks of the paired scores of two variables is worked out for each individual or case, and squared.

(c) The squared differences (D^2) between the paired ranks are totalled for all the cases to give ΣD^2 . The latter is a *measure of disarray* of the ranks of one variable with respect to the

orderly arrangement of the ranks of the other, and have bilaterally symmetric sampling distributions.

(d) In one method, using the n number of pairs of observations in the sample.

$$r_\rho = 1 - \frac{6 \Sigma D^2}{n(n^2 - 1)}.$$

In an alternative method using ΣD_{\max}^2 which is the maximum possible disarray,

$$\Sigma D_{\max}^2 = \frac{n(n^2 - 1)}{3}; \quad r_\rho = 1 - \frac{2 \Sigma D^2}{\Sigma D_{\max}^2}.$$

Inaccuracies :

(i) Inaccuracies of r_ρ result from the use of *average ranks* for tied scores, instead of their true ranks.

(ii) Inaccuracies also arise from *unequal differences* in magnitude between scores given successive ranks. For example, ranks 1, 2, 3 and 4 might have been assigned to scores 10, 12, 17 and 25 respectively so that the differences between the scores bearing consecutive ranks vary widely, amounting to 2 between ranks 1 and 2, 5 between ranks 2 and 3, and 8 between ranks 3 and 4. Because r_ρ is computed from the ranks ignoring such differences in magnitude between successive scores, its value would differ from that of r computed from the same set of scores.

Significance of rho

(a) Where $n \leq 30$, the computed r_ρ is compared directly with the critical r_ρ values for the given n , using a table of critical r_ρ values (Table D of *Appendix*). The computed r_ρ is significant at or below the level whose critical r_ρ is respectively equal to or lower than the computed one ($P \leq \alpha$).

(b) Where $n \geq 10$; a two-tail t test is undertaken to find the probability P of correctness of the H_0 which contends that there is no significant correlation between the

variables. Where s_{r_ρ} is the SE of the computed r_ρ .

$$s_{r_\rho} = \sqrt{\frac{1-r_\rho^2}{n-2}}; \quad t = \frac{r_\rho}{s_{r_\rho}} = r_\rho \sqrt{\frac{n-2}{1-r_\rho^2}};$$

$$df = n - 2.$$

The computed r_ρ is considered significant only if the t score computed from it exceeds or equals the critical t score for the chosen α ($P \leq \alpha$).

(c) Where $n \geq 25$, the computed r_ρ may be alternatively converted to the z score which is referred to unit normal curve areas (Table A) to find the P of the H_0 being correct. The r_ρ is significant only if P is equal to or lower than either the chosen α or the α of 0.05.

$$s_{r_\rho} = \frac{1}{\sqrt{n-1}}; \quad z = \frac{r_\rho}{s_{r_\rho}} = r_\rho \sqrt{n-1};$$

$P = 2 [0.5000 - (\text{area of normal curve from its } \mu \text{ to the computed } z \text{ score})]$.

Example 8.5.1.

Compute r_ρ between heights (cm) and weights (kg) of 12 college students using the following data. Interpret the result.

Student	:	1	2	3	4	5	6	7	8	9	10	11	12
Height	:	165	182	170	162	160	165	170	170	165	176	167	180
Weight	:	58.5	60.0	52.0	48.5	49.5	59.0	49.0	56.0	58.0	60.0	59.5	66.5

Solution :

Table 8.9. Table for computing r_ρ between height and weight.

Student	Height (cm)		Weight (kg)		$D = (R_1 - R_2)$	D^2
	Scores (X)	Ranks (R_1)	Scores (Y)	Ranks (R_2)		
1	165	9	58.5	6	+ 3	9
2	182	1	60.0	2.5	- 1.5	2.25
3	170	5	52.0	9	- 4	16
4	162	11	48.5	12	- 1	1
5	160	12	49.5	10	+ 2	4
6	165	9	59.0	5	+ 4	16
7	170	5	49.0	11	- 6	36
8	170	5	56.0	8	- 3	9
9	165	9	58.0	7	+ 2	4
10	176	3	60.0	2.5	+ 0.5	0.25
11	167	7	59.5	4	+ 3	9
12	180	2	66.5	1	+ 1	1
Total						107.50

1. First method :

(a) Ranks are assigned in descending orders to the scores of height (X) and weight (Y) separately (Table 8.9). Average ranks are given to all the scores of each tied set. Thus, each of three X scores of 170 cm, expected to occupy ranks 4, 5 and 6, gets the average rank of 5 :

$$\text{Average rank} = \frac{\text{Sum of expected ranks}}{\text{Number of tied scores}} = \frac{4+5+6}{3} = 5.$$

The next lower score of 167 consequently gets a rank of 7.

(b) The rank-difference (D) is worked out between the ranks (R_1 and R_2) of the X and Y scores of each pair (Table 8.9).

(c) Each rank-difference is squared and the sum (ΣD^2) of all the squared rank-differences is used in computing r_p .

$$r_p = 1 - \frac{6\Sigma D^2}{n(n^2-1)} = 1 - \frac{6 \times 107.50}{12(144-1)} = +0.62.$$

2. Alternative method :

(a) D and D^2 for all the cases, as well as ΣD^2 , are worked out as in the preceding method (Table 8.9).

(b) ΣD^2_{\max} is computed from n , and used in working out r_p .

$$\Sigma D^2_{\max} = \frac{n(n^2-1)}{3} = \frac{12[(12)^2-1]}{3} = 572.$$

$$r_p = 1 - \frac{2\Sigma D^2}{\Sigma D^2_{\max}} = 1 - \frac{2 \times 107.50}{572} = +0.62.$$

Interpretation :

To find the probability (P) of correctness of the H_0 proposing no correlation, a two-tail t test is undertaken using the r_p computed by any of the preceding methods.

$$s_{r_p} = \sqrt{\frac{1-r_p^2}{n-2}} = \sqrt{\frac{1-(0.62)^2}{12-2}} = 0.248 ; \quad t = \frac{r_p}{s_{r_p}} = \frac{0.62}{0.248} = 2.500 ;$$

$$df = n - 2 = 12 - 2 = 10.$$

Two-tail critical t scores ($df = 10$) are quoted from Table B of Appendix :

$$t_{.01(10)} = 3.169 ; \quad t_{.02(10)} = 2.764 ; \quad t_{.05(10)} = 2.228.$$

As the computed t is higher than the critical $t_{.05}$, the probability P of getting the computed r_p by chances of random sampling is less than 0.05. So, the H_0 can be rejected and the variables have a significant correlation ($P < 0.05$).

Example 8.5.2.

Compute r_p with the data of Example 8.2.3 and interpret the results.

Solution :

1. First method :

(a) Ranks are assigned in descending orders and separately to the vocabulary test scores (X) and the typewriting test scores (Y), arranged in pairs in Table 8.10. Average ranks are given to all the scores of each tied set. Thus, each of two Y scores of 48, expected to occupy ranks 4 and 5, gets the average rank of 4.5 while the next lower score of 41 gets the rank of 6.

Table 8.10. Table for computing r_ρ between vocabulary and typewriting scores.

Student	Vocabulary		Typewriting		$D = (R_1 - R_2)$	D^2
	Scores (X)	Ranks (R_1)	Scores (Y)	Ranks (R_2)		
1	8	9	29	9	0	0
2	22	4	48	4.5	-0.5	0.25
3	35	1	55	1	0	0
4	19	5	49	3	+2	4
5	23	3	53	2	+1	1
6	13	7	41	6	+1	1
7	2	10	22	10	0	0
8	14	6	38	7	-1	1
9	11	8	35	8	0	0
10	25	2	48	4.5	-2.5	6.25
Total						13.50

(b) The rank-difference (D) is worked out between the ranks (R_1 and R_2) of the X and Y scores of each pair.

(c) The rank-differences are squared and all the squared rank-differences are totalled to give the ΣD^2 of 13.50.

(d) r_ρ is computed using ΣD_{\max}^2 and ΣD^2 .

$$\Sigma D_{\max}^2 = \frac{n(n^2 - 1)}{3} = \frac{10(100 - 1)}{3} = 330 ; \quad r_\rho = 1 - \frac{2\Sigma D^2}{\Sigma D_{\max}^2} = 1 - \frac{2 \times 13.50}{330} = +0.92.$$

2. Alternative method :

(a) D and D^2 for all the cases, as well as ΣD^2 , are worked out as in the preceding method (Table 8.10).

(b) r_ρ is next computed using the ΣD^2 thus worked out.

$$r_\rho = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)} = 1 - \frac{6 \times 13.50}{10[(10)^2 - 1]} = +0.92.$$

Interpretation :

Student's t is computed from r_ρ , worked out by either of the above methods, for testing the H_0 of no correlation.

$$s_{r_\rho} = \sqrt{\frac{1 - r_\rho^2}{n - 2}} = \sqrt{\frac{1 - (0.92)^2}{10 - 2}} = 0.139 ; \quad t = \frac{r_\rho}{s_{r_\rho}} = \frac{0.92}{0.139} = 6.619 ;$$

$$df = n - 2 = 10 - 2 = 8.$$

The computed t is compared with two-tail critical t scores ($df = 8$) from Table B.

$$t_{.05(8)} = 2.306 ; \quad t_{.02(8)} = 2.896 ; \quad t_{.01(8)} = 3.355 ; \quad t_{.001(8)} = 5.041.$$

As the computed t is higher than even the critical $t_{.001}$, the computed r_ρ is significant beyond the 0.001 level. So, there is a significant correlation between the variables ($P < 0.001$).

8.6 KENDALL'S RANK CORRELATION

Like Spearman's ρ , the rank correlation coefficient (τ or τ), developed by M.G. Kendall, is a *nonparametric* counterpart of Pearson's r for *simple linear correlation* between two variables. Like r , but in contrast to Pearson's r , τ is computed more easily, has fewer and simpler assumptions, may be applied to small samples, to both normal and non-normal distributions, and to ordinal (ranked) variables as well as such continuous interval and ratio variables whose scores have been changed into ranks ; but it is less powerful than the product-moment r though somewhat more powerful than Spearman's ρ . Values of τ range from -1.00 to $+1.00$, a value of 0.00 indicating the absence of correlation. The values of τ and ρ have high and positive correlations with each other ; but they differ from each other for the same data because of their different scales.

Assumptions

It should be reasonable to assume that :

- (a) there is a *linear relationship* between the variables to be correlated ;
- (b) the magnitudes of scores of each variable can be *expressed in ranks* — so, it is not applicable to nominal variables which cannot be ranked ;
- (c) the paired ranks (scores) of the variables for each individual *occur at random* and *independent* of all other paired ranks (scores) in the sample.

Assumptions for continuous measurement variables and for the normal or near-normal distribution of their scores are not required.

Computation

- (a) Ranks are first assigned in an ascending order of magnitude to the scores of each variable separately, giving average ranks to the scores of each tied set as in the case of

computing r (Table 8.11). But in case of ordinal variables already in ranks in the data, those ranks are directly used in computing τ .

- (b) The ranks (R_1) of the variable with no tied score are serially arranged in an ascending order along a column and each such R_1 rank is paired in the adjoining column with the rank (R_2) of the other variable in the same individual (Table 8.12). If neither or each of the variables has tied scores, the ranks of any of them are arranged as R_1 ranks in the ordered manner pairing them with the ranks (R_2) of the other variable in the respective individuals.

- (c) Starting from the top of the column of the paired R_2 ranks of the second variable, each R_2 rank is taken in turn as the *pivotal rank* and compared with all *subsequent* R_2 ranks in that column. Each subsequent R_2 rank is counted as 1, 0.5 or 0 according as it is higher than, equal to (tied with) or lower than the particular pivotal rank. The count of all the subsequent R_2 ranks for each pivotal rank is entered as the C_i value of the latter (Table 8.12). After using all the R_2 ranks as pivotal ranks in turn and counting their respective C_i values, the latter values are totalled to give ΣC_i .

- (d) To correct the *inaccuracies due to average ranks of tied scores*, correction terms (ΣT_1 and ΣT_2) are computed for the respective variables. In case of each variable, T is first computed for each set of tied scores of that variable, having t number of average ranks in that set : $T = t(t-1)$. The T values of all the tied sets of a variable are then totalled to give ΣT for that variable.

$$\Sigma T_1 = \Sigma [t_1(t_1 - 1)] ;$$

$$\Sigma T_2 = \Sigma [t_2(t_2 - 1)] .$$

If, for example, a variable has two tied sets consisting of 2 and 3 average ranks respectively, $\Sigma T = 2(2-1) + 3(3-1) = 8$. ΣT amounts to 0 if a variable has no tied score.

(e) If, in a sample of size n , both variables have tied scores,

$$\tau = \frac{4\sum C_i - n(n-1)}{\sqrt{[n(n-1) - \sum T_1][n(n-1) - \sum T_2]}}$$

If neither of the variables has any tied score,

$$\tau = \frac{4\sum C_i - n(n-1)}{n(n-1)}$$

2. Alternative method :

After assigning ranks to the scores of the variables and arranging those ranks in pairs along two columns as given in steps (a) and (b) of the preceding method, an alternative method may be followed for the rest of the procedure.

(c) Starting from the top of the column of the paired R_2 ranks, each R_2 rank is in turn taken as the *pivotal rank* and compared with all subsequent R_2 ranks in that column. But each subsequent R_2 rank is counted here as +1, 0 or -1, according as it is higher than, equal to (tied with) or lower than the particular pivotal rank (Table 8.13). The algebraic sum of these counts for each pivotal rank gives the total count C for the latter. After using all the R_2 ranks as pivotal ranks in turn, their C counts are added to give $\sum C$.

(d) Correction terms ($\sum T_1$ and $\sum T_2$) are computed for the tied scores of the respective variables in the same way as in the preceding method.

(e) If both the variables have tied scores,

$$\tau = \frac{2\sum C}{\sqrt{[n(n-1) - \sum T_1][n(n-1) - \sum T_2]}}$$

If neither of the variables has any tied score,

$$\tau = \frac{2\sum C}{n(n-1)}$$

Inaccuracies :

(a) All rank statistics suffer from an inaccuracy owing to the *average ranks* given to tied scores, instead of their true ranks ; this is, however, minimized for *tau* by the correction terms, $\sum T_1$ and $\sum T_2$, in its computation.

(b) Another inaccuracy results from the *unequal differences* between scores given successive ranks.

Significance of tau

(a) For small samples ($n \leq 10$), the computed τ is compared with the critical τ values for the given sample size, taken from a *standard table*. The computed τ is significant at or below the level whose critical τ is either equal to or lower than the computed τ ($P \leq \alpha$).

(b) In case of larger samples ($n > 10$), *tau* has symmetric and near-normal sampling distributions, enabling its conversion to Student's t for testing the H_0 which contends that there is no significant correlation.

$$s_\tau = \sqrt{\frac{2(2n+5)}{9n(n-1)}};$$

$$t = \frac{\tau}{s_\tau} = \tau \sqrt{\frac{9n(n-1)}{2(2n+5)}}; \quad df = \infty.$$

The computed τ is significant at or below that level of significance whose critical t score is either equal to or lower than the computed t ($P \leq \alpha$).

Instead of t , z score may also be computed from the *tau*, using the same formula as for t , and then interpreted with reference to unit normal curve areas (Table A of Appendix).

Example 8.6.1.

Compute τ between pulse rates and respiratory rates per minute, using the following data, to find if there is a significant correlation between the variables.

Individual	:	1	2	3	4	5	6	7	8	9	10	11	12
Respiratory rate	:	15	16	12	21	17	13	18	11	14	20	19	22
Pulse rate	:	72	72	70	82	75	70	72	75	68	84	79	80

Solution :

1. *First method :*

(a) Ranks are assigned separately and in ascending orders to the scores of the two variables. Average rank is given to each score of a tied set (Table 8.11).

(b) Because respiratory rates show no tied score and no average rank, its ranks (R_1) are serially arranged in an ascending order in Table 8.12, pairing each R_1 with the rank R_2 of the pulse rate of the same individual.

(c) As the R_1 ranks include no average rank for tied scores, $\sum T_1$ for R_1 amounts to : $\sum T_1 = \sum [t_1(t_1 - 1)] = 0$. But the R_2 ranks show three tied sets, viz., two ranks of 7.5, two ranks of 2.5 and three ranks of 5,

$$\therefore \sum T_2 = \sum [t_2(t_2 - 1)] = 2(2 - 1) + 2(2 - 1) + 3(3 - 1) = 10.$$

(d) Starting from the top of the column of paired ranks (R_2), each of the latter is taken in turn as the pivotal rank and compared with all subsequent R_2 ranks of that column. Each subsequent R_2 rank is counted as 1, 0.5 or 0, according as it is higher than, equal to (tied with) or lower than the pivotal rank under consideration. The count of all the subsequent ranks for each pivotal R_2 is entered as the C_i of the latter in the table. Thus, for the first pivotal rank 7.5, the subsequent lower ranks 2.5, 2.5, 1, 5, 5 and 5 are counted as 0 each, the subsequent tied rank 7.5 is counted as 0.5, and each of the subsequent higher ranks, 9, 12, 11 and 10, is counted as 1 ; so, the count of ranks (C_i) totals 4.5 for the first pivotal rank 7.5. This is repeated for each R_2 rank in turn, comparing it with the R_2 ranks following it, but not with those preceding it in that column. The C_i values of all the R_2 ranks are then totalled to give the $\sum C_i$ of 51.5.

Table 8.11. Assigning ranks to scores of pulse and respiratory rates.

Individuals	Respiratory rate		Pulse rate	
	Score (X)	Rank (R_1)	Score (Y)	Rank (R_2)
1	15	5	72	5
2	16	6	72	5
3	12	2	70	2.5
4	21	11	82	11
5	17	7	75	7.5
6	13	3	70	2.5
7	18	8	72	5
8	11	1	75	7.5
9	14	4	68	1
10	20	10	84	12
11	19	9	79	9
12	22	12	80	10

Table 8.12. Count of ranks for τ between pulse and respiratory rates (first method).

R_1	R_2	Count of subsequent ranks (C_i)	
1	7.5	0 + 0 + 0 + 0 + 0 + 0.5 + 0 + 1 + 1 + 1 + 1	= 4.5
2	2.5	0.5 + 0 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	= 8.5
3	2.5	0 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	= 8
4	1	1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	= 8
5	5	0.5 + 1 + 0.5 + 1 + 1 + 1 + 1	= 6
6	5	1 + 0.5 + 1 + 1 + 1 + 1	= 5.5
7	7.5	0 + 1 + 1 + 1 + 1	= 4
8	5	1 + 1 + 1 + 1	= 4
9	9	1 + 1 + 1	= 3
10	12	0 + 0	= 0
11	11	0	= 0
12	10	0	= 0
Total			51.5 (ΣC_i)

$$\tau = \frac{4 \Sigma C_i - n(n-1)}{\sqrt{[n(n-1) - \Sigma T_1][n(n-1) - \Sigma T_2]}} = \frac{4 \times 51.5 - 12(12-1)}{\sqrt{[12(12-1) - 0][12(12-1) - 10]}} = + 0.58.$$

2. Alternative method :

(a)-(c) Same as in the previous method, using Tables 8.11 and 8.13.

(d) Each paired rank (R_2) of the second column (Table 8.13) is taken in turn as the pivotal rank and the subsequent ranks of that column are counted as 1, 0 or -1, according as they exceed, equal or are lower than the pivotal rank. These counts are entered in Table 8.13 and totalled for all the pivotal ranks to give the ΣC .Table 8.13. Count of ranks for τ (second method).

R_1	R_2	Count of subsequent ranks (C)	
1	7.5	-1 -1 -1 -1 -1 + 0 -1 + 1 + 1 + 1 + 1	= -2
2	2.5	0 -1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	= 7
3	2.5	-1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	= 7
4	1	1 + 1 + 1 + 1 + 1 + 1 + 1 + 1	= 8
5	5	0 + 1 + 0 + 1 + 1 + 1 + 1	= 5
6	5	1 + 0 + 1 + 1 + 1 + 1	= 5
7	7.5	-1 + 1 + 1 + 1 + 1	= 3
8	5	1 + 1 + 1 + 1	= 4
9	9	1 + 1 + 1	= 3
10	12	-1 -1	= -2
11	11	-1	= -1
12	10	0	= 0
Total			37 (ΣC)

$$\Sigma T_1 = 0 ; \quad \Sigma T_2 = \Sigma [t_2(t_2 - 1)] = 2(2 - 1) + 2(2 - 1) + 3(3 - 1) = 10.$$

$$\tau = \frac{2\Sigma C}{\sqrt{[n(n-1) - \Sigma T_1][n(n-1) - \Sigma T_2]}} = \frac{2 \times 37}{\sqrt{[12(12-1) - 0][12(12-1) - 10]}} = +0.58$$

Interpretation :

As the sample size ($n = 12$) exceeds 10, a two-tail t test is performed using the computed τ , to test the H_0 of no significant correlation.

$$s_\tau = \sqrt{\frac{2(2n+5)}{9n(n-1)}} = \sqrt{\frac{2(2 \times 12 + 5)}{9 \times 12(12-1)}} = 0.221 ; \quad t = \frac{\tau}{s_\tau} = \frac{0.58}{0.221} = 2.624 ; \quad df = \infty.$$

The computed t is compared with the two-tail critical t scores ($df = \infty$) from Table B of Appendix.

$$t_{.02(\infty)} = 2.326 ; \quad t_{.01(\infty)} = 2.576 ; \quad t_{.001(\infty)} = 3.291.$$

The computed t being higher than the critical $t_{.01}$, the probability P of getting the computed τ by chance is too low and amounts to < 0.01 . So, the H_0 is rejected — pulse rate and respiratory rate are considered to have a significant correlation.

Example 8.6.2.

Find whether there is a significant correlation between the test scores of mechanical ability and technical aptitude, using the following data for computing τ .

Student	:	1	2	3	4	5	6	7	8	9	10	11	12
Mechanical ability	:	19	39	22	27	31	28	26	20	21	36	15	9
Technical aptitude	:	6	24	10	13	20	13	10	13	18	14	12	11

Solution :

1. First method :

(a) Ranks are assigned in ascending orders and separately to the scores of the two variables, giving an average rank to each score of a tied set (Table 8.14).

(b) Because mechanical ability scores show no tie and consequently no average rank, their ranks (R_1) are arranged in an ascending order in Table 8.15, pairing each R_1 with the rank R_2 of technical aptitude score of the same individual.

(c) R_1 ranks include no average rank ; so, $\Sigma T_1 = 0$. But R_2 ranks show two tied sets with average ranks, viz., two ranks of 2.5 and three ranks of 7.

$$\therefore \Sigma T_2 = \Sigma [t_2(t_2 - 1)] = 2(2 - 1) + 3(3 - 1) = 8.$$

(d) Starting from the top of the column of paired ranks (R_2), each of the latter is taken in turn as the pivotal rank and compared with all subsequent ranks following it in that column, but not with those preceding it. Each subsequent R_2 rank is counted as +1, 0 or -1 according as it is higher than, equal to (tied with) or lower than the pivotal rank under consideration. The algebraic sum of these counts for each pivotal rank gives the total count C for the latter. For example, for the 4th pivotal R_2 of 7, the subsequent higher ranks of 10, 11, 9 and 12 are each counted as +1, each subsequent tied rank of 7 is counted as 0, and the

subsequent lower ranks of 2.5 and 2.5 are each counted as -1 ; this gives the total (C) of 2 for that pivotal rank (Table 8.15). The C values for all the R_2 ranks are added to give the ΣC of 34.

Table 8.14. Ranking of mechanical ability and technical aptitude scores.

Student	Mechanical ability		Technical aptitude	
	Scores (X)	Rank (R_1)	Scores (Y)	Ranks (R_2)
1	19	3	6	1
2	39	12	24	12
3	22	6	10	2.5
4	27	8	13	7
5	31	10	20	11
6	28	9	13	7
7	26	7	10	2.5
8	20	4	13	7
9	21	5	18	10
10	36	11	14	9
11	15	2	12	5
12	9	1	11	4

Table 8.15. Table for computing *tau* between mechanical ability and technical aptitude scores.

R_1	R_2	Count of subsequent ranks (C)	
1	4	$1 - 1 + 1 + 1 - 1 - 1 + 1 + 1 + 1 + 1 + 1 =$	5
2	5	$- 1 + 1 + 1 - 1 - 1 + 1 + 1 + 1 + 1 + 1 =$	4
3	1	$1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 =$	9
4	7	$1 - 1 - 1 + 0 + 0 + 1 + 1 + 1 =$	2
5	10	$- 1 - 1 - 1 - 1 + 1 - 1 + 1 =$	-3
6	2.5	$0 + 1 + 1 + 1 + 1 + 1 =$	5
7	2.5	$1 + 1 + 1 + 1 + 1 =$	5
8	7	$0 + 1 + 1 + 1 =$	3
9	7	$1 + 1 + 1 =$	3
10	11	$- 1 + 1 =$	0
11	9	1	1
12	12	0	0
Total			34 (ΣC)

$$\tau = \frac{2\Sigma C}{\sqrt{[n(n-1) - \Sigma T_1][n(n-1) - \Sigma T_2]}} = \frac{2 \times 34}{\sqrt{[12(12-1) - 0][12(12-1) - 8]}} = + 0.53.$$

2. Alternative method :

(a)-(c) Same as in the preceding method, using Tables 8.14 and 8.16.

(d) Each paired rank (R_2) of the second column (Table 8.16) is taken in turn as the pivotal rank and the subsequent ranks of that column are counted as 1, 0.5 and 0, according as they are higher than, equal to or lower than the pivotal rank. These counts are entered in Table 8.16 and totalled for all the pivotal ranks to give the ΣC_i .

Table 8.16. Count of ranks for *tau* (second method).

R_1	R_2	Count of subsequent ranks (C_i).
1	4	$1 + 0 + 1 + 1 + 0 + 0 + 1 + 1 + 1 + 1 + 1 = 8$
2	5	$0 + 1 + 1 + 0 + 0 + 1 + 1 + 1 + 1 + 1 = 7$
3	1	$1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 = 9$
4	7	$1 + 0 + 0 + 0.5 + 0.5 + 1 + 1 + 1 = 5$
5	10	$0 + 0 + 0 + 0 + 1 + 0 + 1 = 2$
6	2.5	$0.5 + 1 + 1 + 1 + 1 + 1 = 5.5$
7	2.5	$1 + 1 + 1 + 1 + 1 = 5$
8	7	$0.5 + 1 + 1 + 1 = 3.5$
9	7	$1 + 1 + 1 = 3$
10	11	$0 + 1 = 1$
11	9	$1 = 1$
12	12	$0 = 0$
Total		50 (ΣC_i)

$$\Sigma T_1 = 0 ; \quad \Sigma T_2 = \Sigma [t_2(t_2 - 1)] = 2(2 - 1) + 3(3 - 1) = 8.$$

$$\tau = \frac{4 \Sigma C_i - n(n-1)}{\sqrt{[n(n-1) - \Sigma T_1][n(n-1) - \Sigma T_2]}} = \frac{4 \times 50 - 12(12-1)}{\sqrt{[12(12-1) - 0][12(12-1) - 8]}} = + 0.53.$$

Interpretation :

Because $n > 10$, either z or t may be computed from τ to test the H_0 proposing no significant correlation.

$$s_\tau = \sqrt{\frac{2(2n+5)}{9n(n-1)}} = \sqrt{\frac{2(2 \times 12 + 5)}{9 \times 12(12-1)}} = 0.221 ; \quad z = \frac{\tau}{s_\tau} = \frac{0.53}{0.221} = 2.40.$$

The fractional area of the unit normal curve from μ to the z score of 2.40 amounts to 0.4918 (Table A of Appendix). So, the probability P of the H_0 being correct works out to be very low :

$$P = 2 [0.5000 - (\text{area from } \mu \text{ to } z)] = 2 (0.5000 - 0.4918) = 0.016.$$

So, the two variables have a significant correlation with each other ($P < 0.02$).

[Instead of z , t may also be computed from τ and compared with critical t scores ($df = \infty$) to find the level of significance (see Example 8.6.I)].

8.7 POINT BISERIAL r

This is a specialized form of the product-moment r , used as a measure of linear correlation between a continuous measurement variable and a *dichotomous nominal variable*. It is applicable only when the dichotomous variable is *genuinely dichotomous* with an intervening gap between its two classes; e.g., sex, living-dead, success-failure, right-wrong, HIV positive-HIV negative, pregnant-nonpregnant, Rh⁺-Rh⁻, etc. For example, point biserial r (r_{pbi}) may be computed for correlating either sex or pregnant-nonpregnant states with the serum cholesterol values.

In giving the final form to a *psychological test*, r_{pbi} may be computed to correlate the dichotomously scored right-wrong answers to a given test item with the continuous series of total scores of the entire test; such *item-total correlation* is often undertaken for the selection of test items.

Assumptions

For applying point biserial r , it should be justifiable to assume that :

- one of the variables is a *continuous measurement variable* while the other is a *genuinely dichotomous nominal variable* which cannot yield a continuous or normal distribution even on further exploration ;
- the continuous measurement variable involved has a *normal or near-normal distribution* in the population, without much skewness ;
- in each class of the dichotomous variable, every score of the continuous variable occurs at *random and independent* of all other scores ;
- there is a *linear relation* between the variables.

Computation

Where p and q are the proportions of cases or individuals in the two classes of the dichotomous variable, \bar{X}_p and \bar{X}_q are the mean scores of the continuous variable for individuals or cases belonging to the respective classes of the dichotomous variable, \bar{X} is the grand mean and s_X is the SD of all scores of the continuous variable, and n is the number of individuals in the sample,

$$r_{pbi} = \frac{\bar{X}_p - \bar{X}_q}{s_X} \sqrt{pq} ;$$

$$\text{or, } r_{pbi} = \frac{\bar{X}_p - \bar{X}}{s_X} \sqrt{\frac{p}{q}} .$$

The computation of p and q , may be avoided by using n_p and n_q which are the numbers of cases or individuals in the respective classes :

$$r_{pbi} = \frac{\bar{X}_p - \bar{X}_q}{ns_X} \sqrt{n_p n_q} ;$$

$$\text{or, } r_{pbi} = \frac{\bar{X}_p - \bar{X}}{s_X} \sqrt{\frac{n_p}{n_q}} .$$

The values of r_{pbi} range from -1.00 to $+1.00$, but never reach the extremes. Its value falls with the increasing extremeness of the point of dichotomy of the nominal variable. So, the greater the difference between p and q in the sample, the smaller is the value of r_{pbi} .

Significance

Where the H_0 proposes that there is no correlation between the variables in the population, a *two-tail t test* is done. Where $s_{r_{pbi}}$ is the SE of r_{pbi} ,

$$s_{r_{pbi}} = \sqrt{\frac{1 - r_{pbi}^2}{n - 2}} ; \quad df = n - 2 ;$$

$$t = \frac{r_{pbi}}{s_{r_{pbi}}} = r_{pbi} \sqrt{\frac{n - 2}{1 - r_{pbi}^2}} .$$

Alternatively, for a large sample,

$$s_{r_{pbi}} = \frac{1}{\sqrt{n}}; \quad t = r_{pbi} \sqrt{n}; \quad df = \infty.$$

The computed t is compared with the two-

tail critical t scores. The computed r_{pbi} is considered significant at or below that level of significance whose critical t score is either equal to or lower than the computed t ($P \leq \alpha$).

Example 8.7.1.

The mean vital capacity was found to be 5.32 litres in 77 nonpregnant women and 4.26 litres in 67 pregnant women. The SD of vital capacity for the whole group of 144 women was found to be 1.452 litres. Work out an appropriate correlation coefficient between vital capacity and pregnant-nonpregnant states, and interpret the result.

Solution :

Because vital capacity is a continuous measurement variable and pregnant-nonpregnant states constitute a genuine dichotomous variable, r_{pbi} should be computed.

$$n_p = 77; \quad n_q = 67; \quad n = n_p + n_q = 77 + 67 = 144.$$

$$p = \frac{n_p}{n} = \frac{77}{144} = 0.535; \quad q = \frac{n_q}{n} = \frac{67}{144} = 0.465.$$

$$\bar{X}_p = 5.32 \text{ L}; \quad \bar{X}_q = 4.26 \text{ L}; \quad s_X = 1.452 \text{ L}.$$

$$\therefore r_{pbi} = \frac{\bar{X}_p - \bar{X}_q}{s_X} \sqrt{pq} = \frac{5.32 - 4.26}{1.452} \sqrt{0.535 \times 0.465} = + 0.36.$$

$$[\text{Or, } r_{pbi} = \frac{\bar{X}_p - \bar{X}_q}{ns_X} \sqrt{n_p n_q} = \frac{5.32 - 4.26}{144 \times 1.452} \sqrt{77 \times 67} = + 0.36.]$$

A two-tail t test is done for finding the significance of the computed r_{pbi} . Because the sample is large ($n = 144$),

$$s_{r_{pbi}} = \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{144}} = 0.083. \quad \therefore t = \frac{r_{pbi}}{s_{r_{pbi}}} = \frac{0.36}{0.083} = 4.337; \quad df = \infty.$$

Two-tail critical $t_{.001(\infty)}$ amounts to 3.291 (Table B of Appendix). Because the computed t of 4.337 is higher than even the critical $t_{.001}$, the probability P of correctness of H_0 , proposing no correlation, amounts to less than 0.001 and is considered too low. So, there is a significant correlation between the variables ($P < 0.001$).

Example 8.7.2.

Find if there is a significant linear correlation between sex and pulmonary minute ventilation (L min^{-1}), using the following data of the minute ventilation in 12 men and 8 women.

Men : 6.55, 7.50, 7.16, 9.00, 8.50, 6.24, 7.30, 8.20, 7.45, 8.22, 7.28, 8.00

Women : 6.42, 7.35, 7.05, 6.76, 7.82, 6.60, 7.92, 7.88.

Solution :

Sex is a genuine dichotomous variable while pulmonary ventilation is a continuous measurement variable. So, r_{pbi} is computed.

(a) Proportions of cases (p and q) in the two sexes are worked out.

$$n_p = 12 ; \quad n_q = 8 ; \quad n = n_p + n_q = 12 + 8 = 20.$$

$$p = \frac{n_p}{n} = \frac{12}{20} = 0.6 ; \quad q = \frac{n_q}{n} = \frac{8}{20} = 0.4.$$

(b) Means and over-all SD of the ventilation scores are computed (Table 8.17).

Table 8.17. Table for computing means and SD of pulmonary ventilation scores.

Pulmonary ventilation		$X_p - \bar{X}$	$(X_p - \bar{X})^2$	$X_q - \bar{X}$	$(X_q - \bar{X})^2$
Men (X_p)	Women (X_q)				
6.55	6.42	- 0.91	0.8281	- 1.04	1.0816
7.50	7.35	+ 0.04	0.0016	- 0.11	0.0121
7.16	7.05	- 0.30	0.0900	- 0.41	0.1681
9.00	6.76	+ 1.54	2.3716	- 0.70	0.4900
8.50	7.82	+ 1.04	1.0816	+ 0.36	0.1296
6.24	6.60	- 1.22	1.4884	- 0.86	0.7396
7.30	7.92	- 0.16	0.0256	+ 0.46	0.2116
8.20	7.88	+ 0.74	0.5476	+ 0.42	0.1764
7.45		- 0.01	0.0001		
8.22		+ 0.76	0.5776		
7.28		- 0.18	0.0324		
8.00		+ 0.54	0.2916		
Σ : 91.40	57.80		7.3362		3.0090

$$\bar{X}_p = \frac{\Sigma X_p}{n_p} = \frac{91.40}{12} = 7.62 \text{ L} ; \quad \bar{X} = \frac{\Sigma X_p + \Sigma X_q}{n} = \frac{91.40 + 57.80}{20} = 7.46 \text{ L}.$$

$$s_x = \sqrt{\frac{\Sigma (X_p - \bar{X})^2 + \Sigma (X_q - \bar{X})^2}{n-1}} = \sqrt{\frac{7.3362 + 3.0090}{20-1}} = 0.738 \text{ L}.$$

$$\therefore r_{pbi} = \frac{\bar{X}_p - \bar{X}}{s_x} \sqrt{\frac{p}{q}} = \frac{7.62 - 7.46}{0.738} \sqrt{\frac{0.6}{0.4}} = + 0.266.$$

$$[\text{Or, } r_{pbi} = \frac{\bar{X}_p - \bar{X}}{s_x} \sqrt{\frac{n_p}{n_q}} = \frac{7.62 - 7.46}{0.738} \sqrt{\frac{12}{8}} = + 0.266.]$$

A two-tail t test is undertaken to test the H_0 of no correlation between the variables. Because the sample is small ($n = 20$),

$$s_{r_{pbi}} = \sqrt{\frac{1 - r_{pbi}^2}{n-2}} = \sqrt{\frac{1 - (0.266)^2}{20-2}} = 0.2272 ; \quad t = \frac{r_{pbi}}{s_{r_{pbi}}} = \frac{0.266}{0.2272} = 1.171 ;$$

$$df = n - 2 = 20 - 2 = 18.$$

Two-tail critical t scores ($df = 18$) are quoted from Table B of *Appendix*.

$$t_{.05(18)} = 2.101 ; \quad t_{.10(18)} = 1.734.$$

Because the computed t is lower than even the critical $t_{.10}$, the probability P of the H_0 being correct is too high — there is *no significant correlation* between the variables ($P > 0.10$).

Example 8.7.3.

Using the data presented in the first four columns of Table 8.18, find if there is a significant correlation ($\alpha = 0.05$) between the total test scores of 200 students and their yes/no answers in a dichotomously scored test item in a psychological test.

Solution :

The total test scores constitute a continuous measurement variable while the yes/no answers to the test item constitute a genuine dichotomous variable. So, r_{pti} is computed for the *item-total correlation*.

Table 8.18. Table for computing the means of psychological test scores.

Total test scores	Midpoint (X_c)	Number of students		$f_p X_c$	$f_q X_c$
		'yes' to test item (f_p)	'no' to test item (f_q)		
91 – 105	98	5	0	490	0
76 – 90	83	25	8	2075	664
61 – 75	68	45	20	3060	1360
46 – 60	53	28	25	1484	1325
31 – 45	38	15	10	570	380
16 – 30	23	6	7	138	161
1 – 15	8	2	4	16	32
Total		126 (n_p)	74 (n_q)	7833	3922

(a) Proportions (p and q) of students answering respectively 'yes' and 'no' to the given test item are worked out.

$$p = \frac{n_p}{n_p + n_q} = \frac{126}{126 + 74} = 0.63 ; \quad q = \frac{n_q}{n_p + n_q} = \frac{74}{126 + 74} = 0.37.$$

(b) Mean test scores (\bar{X}_p and \bar{X}_q) of students answering respectively 'yes' and 'no' to the test item and also the grand mean \bar{X} of the total test scores are computed (Table 8.18).

$$\bar{X}_p = \frac{\sum f_p X_c}{n_p} = \frac{7833}{126} = 62.2 ; \quad \bar{X}_q = \frac{\sum f_q X_c}{n_q} = \frac{3922}{74} = 53.0 ;$$

$$\bar{X} = \frac{\sum f_p X_c + \sum f_q X_c}{n_p + n_q} = \frac{7833 + 3922}{126 + 74} = 58.8.$$

Table 8.19. Table for computing the overall SD of total test scores.

Total test scores	X_c	f_p	f_q	$X_c - \bar{X}$	$(X_c - \bar{X})^2$	$f_p(X_c - \bar{X})^2$	$f_q(X_c - \bar{X})^2$
91 - 105	98	5	0	+ 39.2	1536.64	7683.20	0.00
76 - 90	83	25	8	+ 24.2	585.64	14641.00	4685.12
61 - 75	68	45	20	+ 9.2	84.64	3808.80	1692.80
46 - 60	53	28	25	- 5.8	33.64	941.92	841.00
31 - 45	38	15	10	- 20.8	432.64	6489.60	4326.40
16 - 30	23	6	7	- 35.8	1281.64	7689.84	8971.48
1 - 15	8	2	4	- 50.8	2580.64	5161.28	10322.56
Total		126	74			46415.64	30839.36

(c) The over-all SD of total test scores is computed using \bar{X} (Table 8.19).

$$s_X = \sqrt{\frac{\sum f_p(X_c - \bar{X})^2 + \sum f_q(X_c - \bar{X})^2}{n_p + n_q - 1}} = \sqrt{\frac{46415.64 + 30839.36}{126 + 74 - 1}} = 19.70.$$

(d) r_{pbi} is computed as follows.

$$r_{pbi} = \frac{\bar{X}_p - \bar{X}_q}{s_X} \sqrt{pq} = \frac{62.2 - 53.0}{19.70} \sqrt{0.63 \times 0.37} = + 0.23 ;$$

$$[\text{or, } r_{pbi} = \frac{\bar{X}_p - \bar{X}}{s_X} \sqrt{\frac{p}{q}} = \frac{62.2 - 58.8}{19.70} \sqrt{\frac{0.63}{0.37}} = + 0.23]$$

Alternatively,

$$r_{pbi} = \frac{\bar{X}_p - \bar{X}}{s_X} \sqrt{\frac{n_p}{n_q}} = \frac{62.2 - 58.8}{19.70} \sqrt{\frac{126}{74}} = + 0.23.$$

(e) The H_0 contends that there is no significant item-total correlation. A two-tail t test is done to find the P of this H_0 being correct. Because the sample size is large ($n = 200$),

$$t = r_{pbi} \sqrt{n} = 0.23 \sqrt{200} = 3.253 ; \quad df = \infty.$$

$$\alpha = 0.05 ; \quad \text{critical } t_{.05(\infty)} = 1.960.$$

As the computed t exceeds the chosen critical $t_{.05}$, there is a *significant* item-total correlation ($P < 0.05$).

8.8 BISERIAL r

This is a specialized form of the product-moment r , used for *simple linear correlation* between (i) a continuous measurement variable and (ii) either an *apparently dichotomous variable*, seen as consisting of two distinct classes in the data, but expected reasonably to

yield continuous series of metric data on more intensive exploration, or an *artificially dichotomized variable* formed by bisecting the experimentally obtained data of a continuous variable at a point near the median of the latter.

Apparently dichotomous variables include variables such as pass-fail, homeotherm-

poikilotherm, neurotic-non-neurotic, trained-untrained, practised-unpractised and athlete-nonathlete.

On the contrary, the metric scores of a continuous variable may sometimes have either a *truncated distribution* in the sample with no score beyond certain scale values inspite of scores known to occur in the population even beyond such limits, or a *skewed distribution* in the sample inspite of the latter having been drawn from a population having a normal distribution of that variable. The observed distribution of scores of such a variable may be bisected into two classes at a point close to the median of those scores, so as to form an *artificially dichotomized variable* such as diabetic-nondiabetic, normal-hypercholesterolemic and nonhypertensive-hypertensive.

Biserial r may be computed for correlating any of the above two types of dichotomous variables with a continuous measurement variable, e.g., between diabetic-nondiabetic states and serum cholesterol values, hypertension-nohypertension and serum cholesterol concentrations, diabetes-nondiabetes and systolic blood pressure scores, and trained-untrained states of athletes and their vital capacity values.

In psychology, biserial r is frequently used instead of r_{pbi} for (i) an *item-total correlation* between a dichotomized test item and the total scores of a psychological test, on the assumption that the dichotomy of the test item represents an apparent dichotomy of a near-normally distributed variable. Biserial r may also be worked out for (ii) an *item-criterion correlation* between a dichotomized test item and the continuous scores of an external criterion for an attribute, so as to assess the capacity of the test item in measuring that attribute.

The computed values of r_b may cross the

limits of +1.00 and -1.00 if continuous scores of the dichotomized variable are skewed in the population. Moreover, the further away is the point of dichotomy of the dichotomized variable from the median of the continuous scores underlying it, or the smaller the sample size (n), the higher is the SE of r_b and consequently, the lower the dependability of the latter.

Assumptions

For applying biserial r , it should be justifiable to assume that :

(a) one of the variables is a *continuous measurement variable* yielding a continuous metric series of scores while the other is an *apparently or artificially dichotomous variable* which either would yield continuous metric data on further exploration or has been dichotomized from such continuous scores ;

(b) the continuous measurement variable involved has a *normal or near-normal distribution* in the population without much skewness ;

(c) the continuous metric data, underlying the dichotomous variable, have a *unimodal and normal or near-normal distribution* in the population — in case of doubts about this assumption, r_{pbi} should be preferred to r_b ;

(d) the dichotomous variable has been dichotomized at a point *not far from the median* of the continuous distribution of its scores ; in other words, the proportion of cases in each class of the dichotomized variable should not be far different from 0.50 ;

(e) in each class of the dichotomized variable, every score of the continuous variable occurs at *random and independent* of other scores ;

(f) there is a *linear relationship* between the variables.

Computation

Where p and q are the proportions of cases of the sample in two classes of the dichotomized variable, y is the height of the ordinate of the unit normal curve at the point of division of the normal curve area into p and q , \bar{X}_p and \bar{X}_q are the means of the continuous measurement variable for individuals belonging respectively to the classes p and q of the dichotomized variable, \bar{X} is the grand mean and s_X is the overall SD of all the continuous variable scores in the sample,

$$r_b = \frac{\bar{X}_p - \bar{X}_q}{s_X} \times \frac{pq}{y} ; \quad \text{or, } r_b = \frac{\bar{X}_p - \bar{X}}{s_X} \times \frac{p}{y}.$$

The ordinate y is found out from the normal curve table (Table A of Appendix). It is the ordinate of the unit normal curve corresponding to that fractional area from its mean ($\mu = 0$) which equals the absolute difference (neglecting

the algebraic sign) between p and half the area (viz., 0.5000) of the unit normal curve.

If needed, the computed r_b or r_{pbi} may be converted to each other by the following formulae.

$$r_{pbi} = r_b \frac{y}{\sqrt{pq}} ; \quad r_b = r_{pbi} \frac{\sqrt{pq}}{y}.$$

Significance

Where n exceeds 30, neither p nor q is less than 0.10, and the H_0 proposes that there is no correlation between the variables,

$$s_{r_b} = \frac{1}{y} \sqrt{\frac{pq}{n}} ; \quad z = \frac{r_b}{s_{r_b}}.$$

The computed z is then referred to the table of normal curve areas (Table A) for determining the probability P of the H_0 being correct.

Example 8.8.1.

Using the data of Example 8.7.3, compute biserial r for item-total correlation of the relevant psychological test.

Solution :

(a) The following statistics, to be used here, have already been computed in Example 8.7.3.

$$p = 0.63 ; \quad q = 0.37 ; \quad n_p = 126 ; \quad n_q = 74 ; \quad n = 126 + 74 = 200.$$

$$\bar{X}_p = 62.2 ; \quad \bar{X}_q = 53.0 ; \quad \bar{X} = 58.8 ; \quad s_X = 19.70.$$

(b) Since $p = 0.63$, the absolute difference between p and 0.5000 amounts to $|(0.63 - 0.50)| = 0.1300$. Thus, y is the ordinate at the point limiting the area ($p - 0.50$), or 0.1300 from μ in the unit normal curve. By extrapolation from Table A of Appendix, y amounts to 0.3776 for this area of 0.1300.

(c) r_b may be computed by either of the following formulae.

$$\text{Either, } r_b = \frac{\bar{X}_p - \bar{X}_q}{s_X} \times \frac{pq}{y} = \frac{62.2 - 53.0}{19.70} \times \frac{0.63 \times 0.37}{0.3776} = + 0.29,$$

$$\text{or, } r_b = \frac{\bar{X}_p - \bar{X}}{s_X} \times \frac{p}{y} = \frac{62.2 - 58.8}{19.70} \times \frac{0.63}{0.3776} = + 0.29.$$

(d) To test the H_0 that there is no item-total correlation, the computed r_b is converted to z score.

$$s_{r_b} = \frac{1}{y} \sqrt{\frac{pq}{n}} = \frac{1}{0.3776} \sqrt{\frac{0.63 \times 0.37}{200}} = 0.090 ; \quad z = \frac{r_b}{s_{r_b}} = \frac{0.29}{0.09} = 3.22.$$

Two-tail probability P of the H_0 being correct, is given by :

$$P = 2 [0.5000 - (\text{fractional area of unit normal curve from its } \mu \text{ to the } z \text{ score of } 3.22)] \\ = 2 (0.5000 - 0.4994) = 0.0012.$$

As P is too low, there is a *significant item-total correlation* ($P < 0.002$).

Example 8.8.2.

The mean vital capacity amounted to 5.4 litres for 66 athletes and 4.6 litres for 134 nonathletes. The SD of vital capacity was 1.86 litres for the entire sample of 200 cases. Is there any significant correlation between vital capacity and athletic status ?

Solution :

Vital capacity is a continuous measurement variable while athlete-nonathlete status, though a dichotomized variable, is reasonably expected to be a continuous variable in reality. So, r_b is computed in this case.

(a) Proportions of cases (p and q) in two classes of the athlete-nonathlete variable are worked out.

$$n_p = 66 ; \quad n_q = 134 ; \quad n = n_p + n_q = 66 + 134 = 200.$$

$$p = \frac{n_p}{n} = \frac{66}{200} = 0.33 ; \quad q = \frac{n_q}{n} = \frac{134}{200} = 0.67.$$

(b) Since $p = 0.33$, the absolute difference between p and 0.5000 amounts to $|(0.50 - 0.33)| = 0.1700$; so, y is the ordinate at the point limiting the area 0.1700 from μ in the unit normal curve. From Table A, y amounts to 0.3621 for this area of 0.1700.

(c) r_b may be computed using either \bar{X}_p or \bar{X} .

$$\bar{X}_p = 5.4 \text{ L} ; \quad \bar{X}_q = 4.6 \text{ L} ; \quad s_X = 1.86 \text{ L} ; \quad \bar{X} = \frac{n_p \bar{X}_p + n_q \bar{X}_q}{n} = \frac{66 \times 5.4 + 134 \times 4.6}{200} = 4.86 \text{ L}.$$

$$r_b = \frac{\bar{X}_p - \bar{X}_q}{s_X} \times \frac{pq}{y} = \frac{5.4 - 4.6}{1.86} \times \frac{0.33 \times 0.67}{0.3621} = + 0.26,$$

$$\text{or, } r_b = \frac{\bar{X}_p - \bar{X}}{s_X} \times \frac{p}{y} = \frac{5.4 - 4.86}{1.86} \times \frac{0.33}{0.3621} = + 0.26.$$

(d) To test the H_0 that there is no correlation between the variables, the computed r_b is converted to z score.

$$s_{r_b} = \frac{1}{y} \sqrt{\frac{pq}{n}} = \frac{1}{0.3621} \sqrt{\frac{0.33 \times 0.67}{200}} = 0.092 ; \quad z = \frac{r_b}{s_{r_b}} = \frac{0.26}{0.092} = 2.83.$$

$$P = 2 [0.5000 - (\text{fractional area of unit normal curve from } \mu \text{ to the } z \text{ score of } 2.83)] \\ = 2 (0.5000 - 0.4977) = 0.0046.$$

As the P is too low, there is a *significant correlation* between the variables ($P < 0.005$).

8.9 YULE'S PHI COEFFICIENT

Yule's phi coefficient (ϕ) is a nonparametric statistic of correlation between *two genuinely dichotomous nominal variables*, each having two classes with a genuine intervening gap ; e.g., sex, living-dead, success-failure, right-wrong answers, HIV positive-negative, Rh⁺-Rh⁻, pregnant-nonpregnant, etc.

In psychology, ϕ is used for (i) *item-to-item* correlation between two dichotomously scored test items (e.g., scored as yes/no, 1/0 or right/wrong), (ii) *item-criterion correlation* between a dichotomously scored test item and a genuinely dichotomous external criterion representing the attribute under investigation, and (iii) assessment of the *power of a test item* to discriminate between the individuals falling in two classes of the criterion.

Assumptions

For applying ϕ , it should be justifiable to assume that :

(a) both the variables are *genuinely dichotomous* with no reasonable expectation to yield any continuous series of data on more extensive exploration ;

(b) each of the variables has a *bimodal distribution* in the population with a *genuine gap* between its two classes.

Computation

(a) A fourfold (2×2 -fold) *contingency table* is drawn. The top and bottom cells of the right column of the table are named A and C, respectively, while those of the left column are named B and D, respectively (Table 8.20). Classes of one variable are represented along the columns of the table, the right column for the high-value or positive class and the left column for the low-value or negative class. Classes of the other variable are represented along the rows of the table, the top row for the high-value or positive class and the bottom row for the low-value or negative class. Thus, cells B and C contain frequencies of *discordant*

cases, i.e., the frequencies B and C of individuals or cases with positive or high values of one variable and negative or low value of the other. The cells A and D contain frequencies of *concordant cases*, i.e., the frequencies A and D of cases with respectively positive (or high) and negative (or low) values of both variables. A positive correlation is indicated where $AD > BC$, a negative correlation by $BC > AD$, and no correlation if $AD = BC$.

(b) Frequencies of individuals are entered in the cells depending on the high and low value status of the cases with respect to the variables. For example, the frequency of cases, belonging to the high-value class of the column-variable and the low-value class of the row-variable, is entered in the cell C. The total of cell frequencies of each row is entered as the marginal total (f_r) of that row while the total of cell frequencies of each column is entered as the marginal total (f_c) of that column (Table 8.20). Thus, $(A + B)$ and $(C + D)$ give the f_r values while $(A + C)$ and $(B + D)$ give the f_c values. The sample size n equals Σf_r as well as Σf_c .

(c) ϕ is computed as follows.

$$\phi = \frac{AD - BC}{\sqrt{(A+B)(A+C)(B+D)(C+D)}}$$

The computed ϕ ranges from -1.00 to +1.00, but seldom reaches the extremes.

Significance

To test the H_0 proposing no correlation between the variables, the computed ϕ is converted to χ^2 (chi square).

$$\chi^2 = n\phi^2 ; \quad df = 1.$$

The computed χ^2 is compared with the critical χ^2 values ($df = 1$) from Table C of Appendix. The computed ϕ is significant at or below that level of significance which has its critical χ^2 respectively equal to or lower than the computed χ^2 ($P \leq \alpha$).

Example 8.9.1.

Out of 40 pregnant and 60 nonpregnant women, respectively 12 and 24 were found HIV-positive. Is there a significant correlation between pregnancy and HIV-positive test ?

Solution :

(a) The HIV negative-positive variable and the pregnant-nonpregnant variable are represented respectively along columns and rows of a 2×2 -fold contingency table (Table 8.20). The HIV-positive class and the pregnant class are considered the positive classes of the respective variables and represented along the right column and the top row respectively. The frequencies of cases are then entered in the cells according to the classes of the two variables to which each case belongs. The marginal totals, f_r and f_c , are computed for the rows and columns respectively.

Table 8.20. Fourfold contingency table for correlating HIV test results with pregnancy.

	HIV-negative	HIV-positive	Total (f_r)
Pregnant	28 (B)	12 (A)	40 (A + B)
Nonpregnant	36 (D)	24 (C)	60 (C + D)
Total (f_c)	64 (B + D)	36 (A + C)	100 (n)

(b) ϕ is computed using the cell frequencies and marginal totals.

$$\phi = \frac{AD - BC}{\sqrt{(A+B)(A+C)(B+D)(C+D)}} = \frac{12 \times 36 - 28 \times 24}{\sqrt{40 \times 36 \times 64 \times 60}} = -0.10.$$

(c) The computed ϕ is converted to χ^2 for interpretation :

$$\chi^2 = n\phi^2 = 100 \times (-0.10)^2 = 1.00 ; \quad df = 1.$$

The computed χ^2 is found to be lower than even the critical $\chi^2_{30(1)}$ which amounts to 1.07 (Table C of Appendix). So, the computed phi is not significant — there is *no significant correlation* between pregnancy and HIV-test result ($P > 0.30$).

Example 8.9.2.

In a psychological test, 35 persons gave right answers to both of two test items, 25 persons answered both of them wrongly, 20 persons gave right answers to item 2 and wrong answers to item 1, and 10 persons answered item 2 wrongly but item 1 correctly. Is there an item-to-item correlation between the two test items ?

Solution :

(a) The data are arranged in a 2×2 -fold contingency table, recording the right and wrong answers of one item along respectively the top and bottom rows, and those of the other item along respectively the right and left columns (Table 8.21). The frequency of persons, giving a particular combination of right and wrong answers to the two test items, is entered in the specific cell for such combination. For example, the frequency of persons, answering item 1 wrongly but item 2 correctly, is entered in cell C. The marginal totals, f_r and f_c , are worked out for the rows and columns respectively.

Table 8.21. Fourfold contingency table for correlating test items.

Item 1	Item 2		Total (f_r)
	wrong	right	
right	10 (B)	35 (A)	45 (A + B)
wrong	25 (D)	20 (C)	45 (C + D)
Total (f_c)	35 (B + D)	55 (A + C)	90 (n)

(b) ϕ is computed, using the cell frequencies and the marginal totals.

$$\phi = \frac{AD - BC}{\sqrt{(A+B)(A+C)(B+D)(C+D)}} = \frac{35 \times 25 - 10 \times 20}{\sqrt{45 \times 55 \times 35 \times 45}} = + 0.34.$$

(c) The computed ϕ is converted to χ^2 for interpretation :

$$\chi^2 = n\phi^2 = 90 \times (0.34)^2 = 10.40 ; \quad df = 1.$$

Critical χ^2 values : $\chi^2_{.001(1)} = 10.83$, and $\chi^2_{.01(1)} = 6.64$ (vide Table C of Appendix).

The computed χ^2 is higher than the critical $\chi^2_{.01(1)}$; so, the computed ϕ is significant below the 0.01 level. Thus, there is a *significant correlation* between the test items ($P < 0.01$).

8.10 TETRACHORIC r

Tetrachoric r (r_t) is a measure of correlation between two dichotomous variables either (i) when they are only *apparently dichotomous* in the data and may yield normally distributed continuous series of scores on more extensive exploration, or (ii) when their continuous scores in the data have been *artificially dichotomized* into two classes although they have normal distributions in the population (pages 177-178). Such variables include diabetic-nondiabetic, normal-hypertensive, normal-neurotic, homeotherm-poikilotherm, trained-untrained, athlete-nonathlete, practised-unpractised, pass-fail, genetic crossover-noncrossover, etc.

In a psychological test, r_t is often computed during *item analysis* for an *item-criterion correlation* between a dichotomously scored test item and an artificially dichotomized external criterion representing the attribute under study. Thus, r_t may assess the *power of a test item* in exploring the attribute represented by the external criterion.

Assumptions

For computing tetrachoric r , it should be reasonable to assume that :

(a) the variables either have continuous metric data which have been *artificially dichotomized*, or are only *apparently dichotomous* and may yield continuous scores on further exploration ;

(b) such continuous series of scores of the dichotomized variables form *unimodal and normal distributions* in the population ;

(c) the point of dichotomy is *close to the median* of each variable so that neither of the proportions (p and q) of its classes is far from 0.50 ;

(d) there exists a *linear relationship* between the continuous scores of the variables.

Computation

After the data have been arranged in a 2×2 -fold contingency table in the same way as in case of the phi coefficient (§ 8.9), r_t may be

computed approximately by the following *cosine-pie formula*, the original formula being far more complicated. Using the cell frequencies (A , B , C and D) of the contingency table,

$$r_t = \cos \left[\frac{180^\circ \sqrt{BC}}{\sqrt{AD} + \sqrt{BC}} \right]$$

The cosine of the computed angle is worked out using a scientific calculator.

The value of r_t ranges from -1.00 to $+1.00$. When $AD = BC$, the computed angle is 90° so that $r_t = \cos 90^\circ = 0$. When $BC = 0$, the angle amounts to 0° so that $r_t = \cos 0^\circ = +1.00$. When $AD = 0$, the computed angle comes to 180° , making $r_t = \cos 180^\circ = -1.00$. If $AD > BC$, the angle is between 0° and 90° so that r_t lies between 0 and $+1.00$. If $BC > AD$, the angle is between 90° and 180° so that r_t lies between -1.00 and 0 .

If the point of dichotomy is far from the

median, one of the cell frequencies is far lower than the others and the r_t is overestimated. So, r_t should not be computed in case of extreme dichotomies.

The SE (s_{r_t}) of r_t is much higher than that of r . So, for similar stability, the sample size (n) for computing r_t should be nearly double that for computing r . To test the H_0 of no correlation,

$$s_{r_t} = \frac{1}{y_1 y_2} \sqrt{\frac{p_1 p_2 q_1 q_2}{n}} ; \quad t = \frac{r_t}{s_{r_t}} ;$$

where p_1 and q_1 are the proportions of cases in two classes of one variable, p_2 and q_2 are the proportions in the corresponding classes of the other variable, and y_1 and y_2 are the ordinates of the unit normal curve at the points of dichotomy of the respective variables. For finding y_1 and y_2 from p_1 and p_2 , see page 179. The computed t is compared with critical t scores for finding the significance.

Example 8.10.1

49 out of 100 diabetics and 20 out of 120 nondiabetics were hypertensive, the others having normal blood pressure. Is there any significant correlation between diabetes and hypertension?

Solution :

As both the variables are nongenuinely dichotomous, r_t is computed between them.

(a) The data are entered as cell frequencies in a 2×2 -fold contingency table, representing hypertensive cases in the right column and diabetic cases in the top row (Table 8.22).

(b) The cell frequencies are used in computing AD and BC .

$$AD = 49 \times 100 = 4900 ; \quad BC = 51 \times 20 = 1020 ; \quad \therefore AD > BC.$$

(c) The cosine-pie formula is used in computing r_t .

$$r_t = \cos \left[\frac{180^\circ \sqrt{BC}}{\sqrt{AD} + \sqrt{BC}} \right] = \cos \left[\frac{180^\circ \sqrt{1020}}{\sqrt{4900} + \sqrt{1020}} \right] = \cos 56.39^\circ = +0.55.$$

(d) The computed r_t is converted to Student's t for testing the H_0 of no correlation. Where p_1 and q_1 are the respective proportions of diabetic and nondiabetic cases, p_2 and q_2 are those of hypertensive and nonhypertensive cases, and n is the sample size,

$$p_1 = \frac{100}{220} = 0.455 ; \quad q_1 = \frac{120}{220} = 0.545 ; \quad p_2 = \frac{69}{220} = 0.314 ; \quad q_2 = \frac{151}{220} = 0.686.$$

Table 8.22. Fourfold contingency table for correlating diabetes and hypertension.

	Nonhypertensive	Hypertensive	Total (f_r)
Diabetic	51 (B)	49 (A)	100
Nondiabetic	100 (D)	20 (C)	120
Total (f_c)	151	69	220 (n)

The unit normal curve ordinates, y_1 and y_2 at the points of dichotomy of the respective variables, are found out from the differences between half the unit normal curve area, viz., 0.5000, and p_1 and p_2 respectively.

$$0.500 - p_1 = 0.500 - 0.455 = 0.045 ; \quad y_1 = 0.396 \text{ (Table A).}$$

$$0.500 - p_2 = 0.500 - 0.314 = 0.186 ; \quad y_2 = 0.355 \text{ (Table A).}$$

$$s_{r_1} = \frac{1}{y_1 y_2} \sqrt{\frac{p_1 p_2 q_1 q_2}{n}} = \frac{1}{0.396 \times 0.355} \sqrt{\frac{0.455 \times 0.314 \times 0.545 \times 0.686}{220}} = 0.111.$$

$$t = \frac{r_1}{s_{r_1}} = \frac{0.55}{0.111} = 4.955 ; \quad df = \infty.$$

Critical t scores ($df = \infty$) are quoted from Table B.

$$t_{.05(\infty)} = 1.960 ; \quad t_{.01(\infty)} = 2.576 ; \quad t_{.001(\infty)} = 3.291.$$

Because the computed t of 4.955 is higher than even $t_{.001}$, there is a *significant correlation* between diabetes and hypertension ($P < 0.001$).

8.11 CONTINGENCY COEFFICIENT

Contingency coefficient (C) is a *nonparametric correlation coefficient* between two variables, either or both of which are divided into more than two classes with intervening gaps.

Assumptions

No assumption is needed for the genuineness of the discrete or discontinuous nature of distributions of the variables ; it is applicable both to variables with genuine gaps between their classes, and to those which either may be resolved into continuous data on further exploration or have been dichotomized on the basis of their continuous distributions.

Nor is the assumption needed for normality or near-normality of their distributions. It should, however, be reasonable to assume that each individual or case occurs in the sample *at random* and *independent* of all others.

Computation

Contingency coefficient is computed from the observed frequencies (f_o) of the data and the expected frequencies (f_e) calculated on the basis of the H_0 which proposes that there is no significant correlation between the variables.

(a) A contingency table is framed with the classes of one variable represented along the columns and the classes of the other variable along the rows (Table 8.23).

(b) The frequencies (f_o) of cases are entered in the respective cells of the table according to the classes of the two variables to which those cases belong. The marginal total of each row (f_r) and that of each column (f_c) are also entered in the relevant marginal column and row, respectively.

(c) Where r and c are the numbers of respectively rows and columns containing cells with f_o values, $df = (r-1)(c-1)$.

(d) Choosing at random as many cells as the computed df , the f_e corresponding to the f_o of each such cell is worked out as follows :

$$f_e = \frac{f_r f_c}{n}$$

where f_r and f_c are the marginal totals of respectively the row and the column of the relevant cell, and n is the total frequency of cases.

The f_e of each of the other cells is then computed by subtracting the sum of f_e values, already worked out in that row (or column), from the relevant marginal total (f_r or f_c).

(e) The sum of quotients (S) and the contingency coefficient (C) are then computed using the f_o and f_e of each cell.

$$S = \sum \frac{f_o^2}{f_e} ; \quad C = \sqrt{1 - \frac{n}{S}}$$

The value of C may range upto 1.00, but does not reach even 0.90 unless the contingency table is higher than 5×5 -fold. C rises with the number (k) of classes of the variables, but suffers from errors if k is large, because some f_o values may then fall below 5.

Significance

The H_0 proposes that there is no correlation between the variables. The computed C is converted to χ^2 to test this H_0 .

$$\chi^2 = \frac{nC^2}{1-C^2} ; \quad df = (r-1)(c-1)$$

The computed χ^2 is compared with critical χ^2 values for different levels of significance. The contingency coefficient may be considered significant at or below that level whose critical χ^2 either equals or falls below the computed χ^2 ($P \leq \alpha$).

Example 8.11.1.

Out of 60 people of Mongolian race, 14, 20, 17 and 9 had blood groups of respectively O, A, B and AB. But of 40 persons of Polynesian race, respectively 6, 10, 15 and 9 individuals belonged to those blood groups. Is there a significant correlation between race and blood group ?

Solution :

Because race and blood group are divided into 2 and 4 classes respectively, C is computed for their correlation.

(a) The data are arranged in a 4×2 -fold contingency table (Table 8.23). Marginal totals (f_r and f_c) of the rows and columns are computed.

$$df = (r-1)(c-1) = (2-1)(4-1) = 3 ; \quad n = 60 + 40 = 100.$$

(b) The df being 3, the f_e corresponding to the f_o of any 3 cells, chosen at random, is computed. For example, for the cell corresponding to the Mongolian race and the A group,

$$f_e = \frac{f_r f_c}{n} = \frac{60 \times 30}{100} = 18.$$

Table 8.23. Table for computing contingency coefficient between race and blood group.

	O group		A group		B group		AB group		Total (f_r)
	f_o	f_e	f_o	f_e	f_o	f_e	f_o	f_e	
Mongolian	14	12	20	18	17	19	9	11	60
Polynesian	6	8	10	12	15	13	9	7	40
Total (f_c)	20	20	30	30	32	32	18	18	100 (n)

Similarly, for the cell corresponding to the Polynesian race and the blood group O,

$$f_e = \frac{f_r f_c}{n} = \frac{40 \times 20}{100} = 8;$$

and for the cell corresponding to the Polynesian race and the B group,

$$f_e = \frac{f_r f_c}{n} = \frac{40 \times 32}{100} = 13.$$

The f_e of each of the other cells is computed by subtracting the already computed f_e values from the relevant f_r or f_c value. For example, the f_e amounting to 18 for the Mongolian vs A group cell is deducted from the f_c of 30 for the A group to give the f_e amounting to 12 for the Polynesian vs A group cell.

(c) The statistics S and C are then computed, using f_o^2 and f_e values of each cell.

$$S = \sum \frac{f_o^2}{f_e} = \frac{(14)^2}{12} + \frac{(20)^2}{18} + \frac{(17)^2}{19} + \frac{(9)^2}{11} + \frac{(6)^2}{8} + \frac{(10)^2}{12} + \frac{(15)^2}{13} + \frac{(9)^2}{7} = 102.84.$$

$$C = \sqrt{1 - \frac{n}{S}} = \sqrt{1 - \frac{100}{102.84}} = 0.166.$$

(d) C is converted to χ^2 for testing its significance.

$$\chi^2 = \frac{nC^2}{1 - C^2} = \frac{100(0.166)^2}{1 - (0.166)^2} = 2.83; \quad df = (r - 1)(c - 1) = (2 - 1)(4 - 1) = 3.$$

From Table C of Appendix, critical $\chi_{.05(3)}^2 = 7.82$, $\chi_{.30(3)}^2 = 3.66$, and $\chi_{.20(3)}^2 = 4.64$. Thus, the computed χ^2 of 2.83 is lower than the critical $\chi_{.05}^2$ and even the critical $\chi_{.30}^2$. So, the probability P of getting the computed χ^2 and hence the computed C by chance, exceeds 0.30. Hence, there is *no significant correlation* between blood groups and races ($P > 0.30$).

8.12 REGRESSION

Regression is prediction statistics. It predicts the most likely value of a variable on the basis of the given value(s) of another or other variable(s). The variable, whose values are predicted, is the *dependent variable* or *criterion*; the variable whose values form the basis of the prediction is called the *independent variable* or *predictor*. Regression can be worked out only if the dependent variable and

the independent variable(s) possess *significant correlation* with each other. It translates the relation between two or more variables into an expression showing one of them as a function of the other(s). Regression, like correlation, holds good only in a particular population to which the sample belongs, and only for that limited range of scores of the variables from which it has been derived; it cannot be extended beyond these limits.

Types of regression

Regression may be either *simple* or *multiple* according to the number of predictors involved.

(a) In *simple regression*, the criterion or dependent variable is a function of a single independent variable or predictor — the scores of the former are predicted from the given scores of the single predictor ; e.g., the regression of height of a person on his/her weight ; the regression of examination marks of a candidate in Mathematics on his/her numerical aptitude test score ; the regression of oxygen consumption on the tracheal ventilation of an insect. (b) In *multiple regression*, the criterion is a function of two or more predictors ; thus, its scores are predicted from the scores of more than one predictor ; e.g., the regression of surface area of a person on his/her height and weight ; the regression of Maths marks of an examinee on his/her numerical aptitude and abstract reasoning test scores.

Regression may again be either *linear* or *nonlinear* according as the relation between the criterion and the predictor can be described in terms of a straight line or a curved line. (a) *Linear regression* expresses the dependent variable Y as the linear function of the independent variable X . In other words, the scattergram of the scores of criterion, plotted against those of predictor, should show a linear distribution of its plotted points. (b) For a *nonlinear regression*, the scattergram should show a curvilinear distribution like elliptical and hyperbolic ones.

Models of regression

There are three models of regression according to the nature(s) of the independent variable(s).

(a) Model I regression :

It is the regression of a dependent variable or criterion (Y) on an independent variable or predictor (X) which is a 'fixed' treatment

variable (page 5). A model I regression is used in predicting the values of Y for specific values of X when the latter is varied by the investigator at precise and predetermined manners and rates. The values of Y suffer from errors due to random variations. But the values of the 'fixed' treatment variable X are free from random errors, because they vary under the planned and deliberate control of the investigator and not at random. So, model I regression can estimate how much of the variations of Y may result from the variations of X and can thus explore the causation of the changes in the dependent variable Y due to the changes in the independent variable X , i.e., their *cause-and-effect relation*. In a *simple model I regression*, the lone predictor is a 'fixed' treatment variable ; e.g., the regression of tracheal ventilation of an insect on 'fixed' and predetermined doses of an insecticide ; the regression of blood sugar level on predetermined doses of injected insulin. In a *multiple model I regression*, based on a combination of two predictors, both of the latter are 'fixed' treatments.

(b) Model II regression :

It is the regression of a criterion (Y) on a predictor (X) which is a *classification variable* beyond the control of the investigator (page 5). It predicts the most likely value of Y on the basis of an already existing value of X in the individual — X is measured, but not applied, by the investigator. Because the values of the predictor are not 'fixed', controlled or applied by the investigator, its values suffer from random errors. Thus, both the dependent and independent variables vary at random and have random errors. So, model II regression cannot explore the cause-and-effect relation between the variables. It merely expresses an *interdependence of their changes*, sometimes due to common causes. The regressions of cardiac stroke volume on venous return, of

glomerular filtration rate on glomerular blood pressure, of tracheal ventilation of an insect on atmospheric temperature, and of examination marks in a language on the verbal ability test score of an examinee are examples of *simple model II regression*; each uses a random predictor such as venous return, glomerular blood pressure, atmospheric temperature, and verbal ability. The Dubois-Dubois formula for human body surface area is a *multiple model II regression* of surface area on height and weight of the individual.

(c) *Model III regression* :

This is always a *multiple regression* predicting the likely value of a dependent variable from the given values of two or more predictors, some of the latter being 'fixed' treatment(s) and the other(s) classification variable(s); e.g., the regression of blood thyroxine level on combinations of atmospheric temperature (classification variable) and injected dose of thyrotropin ('fixed' treatment).

Properties of simple linear regression

1. Linear regression of a variable Y on the basis of scores of another variable X , or vice versa, can be worked out only when the two variables have a *significant linear correlation*, r_{YX} or r_{XY} .

2. The *scattergram*, resulting from the plotting of the predicted criterion scores (\hat{Y}) against the corresponding predictor scores (X) used in their regression, has a linear distribution with an upward or downward slope (page 34, Fig. 2.13).

3. The \hat{Y} values, predicted from those of X , lie around a straight line called the *regression line of Y on X* . The sum of squared vertical distances of the points, plotted with the paired scores of the variables, from this line is kept at a minimum (*method of least squares*). In other words, it is the *best-fitting straight line* for

those plotted points.

4. Two separate regression equations may be worked out for each pair of variables (X and Y). One of these is the regression equation of Y on X , predicting the Y scores on the basis of the X scores; the other is the regression equation of X on Y , predicting the X scores on the basis of the Y scores. However, the regression of that one of two correlated variables is generally worked out, which is relatively more difficult to measure or is measured less precisely than the other. Thus, it is more practical to work out and use the regression of oxygen consumption on vital capacity than the other way around, because the latter is easier and less time-consuming to measure than the former.

5. The *regression equation of Y on X* is given by the following :

$$\hat{Y} = a_{YX} + b_{YX} X,$$

$$\text{or, } \hat{Y} = \bar{Y} + b_{YX} (X - \bar{X}),$$

where \hat{Y} is the Y score predicted in an individual on the basis of the actually measured X score of the latter, b_{YX} is the slope and a_{YX} is the general level of the regression line showing Y as a linear function of X .

6. The statistic b_{YX} is a prediction statistic and is called the *regression coefficient* of Y on X . It is the average rate of increase or decrease in the score of the criterion Y for unit rise in the score of the predictor X . It is given basically by the ratio of the covariance of scores of both variables and the variance of scores of the predictor. Thus,

$$b_{YX} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2};$$

$$\text{or, } b_{YX} = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2};$$

$$\text{or, } b_{YX} = r_{YX} \frac{s_Y}{s_X}.$$

7. The statistic a_{YX} is the *Y-intercept* of the regression line of Y on X .

$$a_{YX} = \bar{Y} - b_{YX} \bar{X}.$$

8. Likewise, the regression equation of X on Y may also be worked out. Where \hat{X} is the X score predicted on the basis of a given Y score, b_{XY} is the regression coefficient of X on Y and gives the slope of the regression line, and a_{XY} gives the general level of that line,

$$b_{XY} = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(Y - \bar{Y})^2};$$

$$\text{or, } b_{XY} = \frac{n\sum XY - \sum X \sum Y}{n\sum Y^2 - (\sum Y)^2}$$

$$\text{or, } b_{XY} = r_{XY} \frac{s_X}{s_Y}, \quad (r_{XY} = r_{YX}).$$

$$a_{XY} = \bar{X} - b_{XY} \bar{Y}.$$

$$\hat{X} = a_{XY} + b_{XY} Y,$$

$$\text{or, } \hat{X} = \bar{X} + b_{XY} (Y - \bar{Y}).$$

9. These two regression equations as well as the two regression lines become identical with each other only when the product-moment r (viz., r_{XY} or r_{YX}) equals either +1.00 or -1.00. For all other values of r , the regression lines intersect at a point which corresponds to the means (\bar{X} and \bar{Y}) of the two variables.

10. The angle between the two regression lines increases with the decrease in the magnitude of r and reaches 90° when r amounts to 0.00 — both b_{YX} and b_{XY} become 0 in value in such a case and no regression is possible.

11. The geometric mean of the two regression coefficients equals the product-moment r between the two variables.

$$r_{YX} \text{ or } r_{XY} = \sqrt{b_{YX} b_{XY}}.$$

It follows that the two regression coefficients are reciprocals of each other when

the product-moment r between the variables amounts to either +1.00 or -1.00.

12. The predicted score of the criterion (say, \hat{Y}) is merely its *most probable score* in an individual having a given score (say, X) of the predictor. So, in any sample, all the individuals with a given X score are not expected to have their actual Y scores identical with the \hat{Y} score predicted by regression on the given X score. Rather, the actual Y scores, experimentally measured in all such individuals, will form a normal or near-normal distribution around the relevant \hat{Y} score with the latter as the mean.

13. For all values of the predictor score X , the experimentally measured Y scores of the criterion should be scattered to similar extents around the regression line. In other words, the deviations of observed Y scores from the respective predicted \hat{Y} scores should obey *homoscedasticity* all along the regression line.

14. The difference between the predicted \hat{Y} score and the observed Y scores of the criterion is called the *error of prediction*. It rises with the decrease in the magnitude of correlation and is measured as the *SE of estimate* (s_{YX}) of Y on X .

$$s_{YX} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n}} = s_Y \sqrt{1 - r_{YX}^2}.$$

Evidently, when r_{YX} amounts to +1.00 or -1.00, s_{YX} amounts to 0 and all the observed Y scores coincide with the respective \hat{Y} scores so as to lie right on the regression line. On the contrary, when r_{YX} amounts to 0.00, s_{YX} equals s_Y and no prediction is possible.

Similarly, the *SE of estimate* (s_{XY}) of X on Y is given by :

$$s_{XY} = \sqrt{\frac{\sum(X - \hat{X})^2}{n}} = s_X \sqrt{1 - r_{XY}^2}.$$

Assumptions for simple linear regression

It should be reasonable to assume that :

(a) the variables involved in regression are either *continuous measurement variables* or such apparently discrete variables as can be reasonably expected to yield *continuous metric scores* on further exploration ;

(b) both the variables have *unimodal and fairly symmetrical distributions* in the population ;

(c) the scores (Y) of the dependent variable (criterion) is a *linear function* of the scores (X) of the independent variable (predictor) — in other words, there is a *significant product-moment r* between the variables and their scattergram has a linear distribution ;

(d) the Y scores of the criterion, measured in a large number of individuals possessing a given X score of the predictor, are *distributed normally*, independent of each other and around the predicted \hat{Y} score as the mean, and the variances of all such distributions around the respective \hat{Y} scores obey *homoscedasticity* ;

(e) the predictor variable is either a “fixed” experimental treatment in model I regression, or a classification variable in model II regression.

Computation of simple linear regression

The *regression coefficient* b_{YX} for the regression of Y on X is computed by using any of the following formulae.

1. Using raw scores :

(a) The X and Y scores of the variables are totalled separately to give ΣX and ΣY respectively.

(b) Each X score is squared and all these squared X scores are totalled to give ΣX^2 .

(c) Each X score is multiplied by the Y score of the same individual to get XY of that

individual, and all such XY values are totalled to give ΣXY .

(d) The statistic b_{YX} is computed as follows using the sample size n (Example 8.12.1).

$$b_{YX} = \frac{n\Sigma XY - \Sigma X\Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}.$$

2. Using the sum of products :

(a) The means (\bar{X} and \bar{Y}) of the variables are computed using ΣX and ΣY .

$$\bar{X} = \frac{\Sigma X}{n} ; \quad \bar{Y} = \frac{\Sigma Y}{n}.$$

(b) The deviations of X and Y scores from their respective means are computed for each individual and then multiplied with each other to give $(X - \bar{X})(Y - \bar{Y})$. These products are added up to give the *sum of products*, $\Sigma(X - \bar{X})(Y - \bar{Y})$.

(c) The deviations of X scores from \bar{X} are squared and the squared deviations are totalled to give the *sum of squares* of the predictor, viz., $\Sigma(X - \bar{X})^2$.

(d) The regression coefficient is then computed as follows (Example 8.12.2).

$$b_{YX} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2}.$$

3. Using product-moment r :

b_{YX} may be computed using the product-moment r (r_{YX}) and the *SDs* (s_X and s_Y) of the variables (Example 8.12.3).

$$b_{YX} = r_{YX} \frac{s_Y}{s_X}.$$

4. Using covariance and variance :

(a) ΣX , ΣY , ΣX^2 and ΣXY are worked out as in the steps (a) to (c) of the computation from raw scores.

(b) $Cov(X, Y)$ and variance (s_X^2) of X scores are then computed using the sample size n .

$$\text{Cov}(X, Y) = \frac{\sum XY}{n} - \frac{\sum X \sum Y}{n^2};$$

$$s_X^2 = \frac{\sum X^2}{n} - \frac{(\sum X)^2}{n^2}$$

(c) b_{YX} is then computed as follows (Example 8.12.1).

$$b_{YX} = \frac{\text{Cov}(X, Y)}{s_X^2}.$$

The b_{YX} computed by any of the preceding alternative methods is then used in working out a_{YX} .

$$a_{YX} = \bar{Y} - b_{YX} \bar{X}.$$

The computed scores of a_{YX} and b_{YX} are then put in the following equation to give the regression equation of Y on X .

$$\hat{Y} = a_{YX} + b_{YX} X.$$

To draw the linear regression line, a number of X scores are chosen at random from within the range of X in the sample. The \hat{Y} score for each of them is computed using the regression equation. Each \hat{Y} score is plotted graphically against the corresponding X score (Fig. 8.4).

Example 8.12.1.

Work out the linear regression equation of height (cm) on weight (kg), using the following data from 10 college students.

Student :	1	2	3	4	5	6	7	8	9	10
Height :	165	182	170	162	160	165	170	165	176	180
Weight :	58.5	64	52	48.5	49.5	59	59	58	60	67

Solution :

1. First method using raw scores :

- The height (Y) and the weight (X) scores are totalled to give $\sum X$ and $\sum Y$ respectively (Table 8.24).
- Each Y score is multiplied by the corresponding X score and all these products are totalled to give $\sum XY$.
- Each X score is squared and the X^2 values are totalled to give $\sum X^2$ (Table 8.24).

$$b_{YX} = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} = \frac{10 \times 97872.5 - 575.5 \times 1695}{10 \times 33439.75 - (575.5)^2} = 1.017.$$

2. Alternative method using covariance :

- (a)-(c) same as in the preceding method.
- (d) The covariance of X and Y scores and the variance of X scores are worked out and used in computing b_{YX} .

$$\text{Cov}(X, Y) = \frac{\sum XY}{n} - \frac{\sum X \sum Y}{n^2} = \frac{97872.5}{10} - \frac{575.5 \times 1695}{100} = 32.525.$$

$$s_X^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n} \right)^2 = \frac{33439.75}{10} - \left(\frac{575.5}{10} \right)^2 = 31.9725.$$

$$b_{YX} = \frac{\text{Cov}(X, Y)}{s_X^2} = \frac{32.525}{31.9725} = 1.017.$$

3. Computation of a_{YX} and the regression equation :

(a) a_{YX} is computed using the b_{YX} worked out by any of the preceding methods.

$$\bar{Y} = \frac{\sum Y}{n} = \frac{1695}{10} = 169.5 ; \quad \bar{X} = \frac{\sum X}{n} = \frac{575.5}{10} = 57.55.$$

$$a_{YX} = \bar{Y} - b_{YX} \bar{X} = 169.5 - 1.017 \times 57.55 = 110.97.$$

(b) The values of a_{YX} and b_{YX} are used to give the regression equation.

$$\hat{Y} = a_{YX} + b_{YX} X ; \quad \text{or, } \hat{Y} = 110.97 + 1.02 X.$$

Table 8.24. Table for computing regression coefficient from raw scores/covariance.

Student	Height (Y)	Weight (X)	XY	X ²
1	165	58.5	9652.5	3422.25
2	182	64	11648	4096
3	170	52	8840	2704
4	162	48.5	7857	2352.25
5	160	49.5	7920	2450.25
6	165	59	9735	3481
7	170	59	10030	3481
8	165	58	9570	3364
9	176	60	10560	3600
10	180	67	12060	4489
Σ	1695	575.5	97872.5	33439.75

Example 8.12.2.

Work out a linear regression equation of typewriting score on vocabulary test score using the following data.

Individual	:	1	2	3	4	5	6	7	8	9
Vocabulary score	:	8	22	35	19	23	13	2	14	11
Typewriting score	:	29	48	55	45	50	35	18	38	30

Solution :

1. First method using the sum of products :

(a) The means of vocabulary scores (X) and typewriting scores (Y), the deviations of X and Y scores from their respective means, the sum of products of the deviations of X and Y scores from their respective means, and the sum of squared deviations of X scores from \bar{X} are worked out (Table 8.25).

Table 8.25. Table for computing regression coefficient from the sum of products.

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})^2$	$(X - \bar{X})(Y - \bar{Y})$
				68.89	80.51
8	29	- 8.3	- 9.7	32.49	53.01
22	48	+ 5.7	+ 9.3	349.69	304.81
35	55	+18.7	+16.3	7.29	17.01
19	45	+ 2.7	+ 6.3	44.89	75.71
23	50	+ 6.7	+11.3	10.89	12.21
13	35	- 3.3	- 3.7	204.49	296.01
2	18	-14.3	-20.7	5.29	1.61
14	38	- 2.3	- 0.7	28.09	46.11
11	30	- 5.3	- 8.7		
Σ	147	348		752.01	886.99

$$\bar{X} = \frac{\Sigma X}{n} = \frac{147}{9} = 16.3 ; \quad \bar{Y} = \frac{\Sigma Y}{n} = \frac{348}{9} = 38.7.$$

$$\Sigma(X - \bar{X})^2 = 752.01 ; \quad \Sigma(X - \bar{X})(Y - \bar{Y}) = 886.99.$$

(b) The regression coefficient b_{YX} is computed using the sum of products and the sum of squares of X .

$$b_{YX} = \frac{\text{Sum of products}}{\text{Sum of squares of } X} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2} = \frac{886.99}{752.01} = 1.18.$$

2. Alternative method using raw scores :

(a) The X and Y scores are totalled to give ΣX and ΣY respectively (Table 8.26).

(b) Each Y score is multiplied by the corresponding X score and all such products are totalled to give ΣXY .

(c) Each X score is squared and the X^2 values are totalled to give ΣX^2 (Table 8.26).

Table 8.26. Table for computing regression coefficient from raw scores.

X	X^2	Y	XY
8	64	29	232
22	484	48	1056
35	1225	55	1925
19	361	45	855
23	529	50	1150
13	169	35	455
2	4	18	36
14	196	38	532
11	121	30	330
Σ	147	3153	348
			6571

(d) The data of Table 8.26 are used in working out b_{YX} .

$$b_{YX} = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{9 \times 6571 - 147 \times 348}{9 \times 3153 - (147)^2} = 1.18.$$

3. Computation of a_{YX} and the regression equation :

$$\bar{X} = \frac{\sum X}{n} = \frac{147}{9} = 16.3 ; \quad \bar{Y} = \frac{\sum Y}{n} = \frac{348}{9} = 38.7 ;$$

$$a_{YX} = \bar{Y} - b_{YX}\bar{X} = 38.7 - 1.18 \times 16.3 = 19.47 ;$$

$$\hat{Y} = a_{YX} + b_{YX}X ; \quad \text{or, } \hat{Y} = 19.47 + 1.18X.$$

Example 8.12.3.

Compute the regression equation of Maths marks on the differential aptitude test (DAT) scores, using the following data, and work out the SE of estimate.

No. of students = 64. Mean DAT score (\bar{X}) = 90.2 ($s_X = 28.35$).
Mean Maths marks (\bar{Y}) = 52.5 ($s_Y = 22.40$). $r_{YX} = +0.60$.

Solution :

$$r_{YX} = +0.60 ; \quad n = 64 ; \quad \bar{X} = 90.2 ;$$

$$s_X = 28.35 ; \quad \bar{Y} = 52.5 ; \quad s_Y = 22.40.$$

$$\therefore b_{YX} = r_{YX} \frac{s_Y}{s_X} = 0.60 \times \frac{22.40}{28.35} = 0.47 ;$$

$$a_{YX} = \bar{Y} - b_{YX}\bar{X} = 52.5 - 0.47 \times 90.2 = 10.11.$$

$$\therefore \hat{Y} = a_{YX} + b_{YX}X = 10.11 + 0.47X.$$

$$s_{YX} = s_Y \sqrt{1 - r_{YX}^2} = 22.40 \sqrt{1 - (0.60)^2} = 17.92.$$

Two \hat{Y} scores are computed, using two X scores chosen at random from within the range of X scores, and are plotted against the respective X scores (Fig. 8.4).

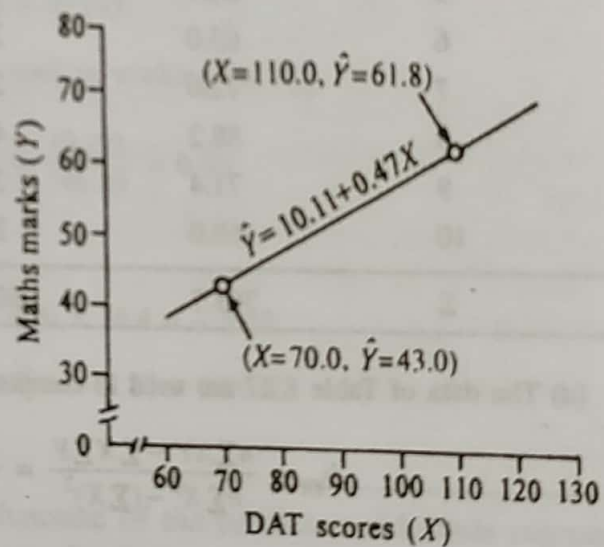


Fig. 8.4. Regression line of Maths marks on DAT scores.

$$(a) \text{ Where } X = 70, \quad \hat{Y} = 10.11 + 0.47 \times 70 = 43.0$$

$$(b) \text{ Where } X = 110, \quad \hat{Y} = 10.11 + 0.47 \times 110 = 61.8.$$

Example 8.12.4.

Work out the linear regression equation of oxygen consumption (Y ml per minute) on tracheal ventilation (X ml per minute) using the following data of a sample of ten insects.

Individual :	1	2	3	4	5	6	7	8	9	10
Y ml :	3.0	4.0	3.3	2.8	3.5	2.2	2.5	4.2	2.3	2.3
X ml :	75.2	86.0	78.0	75.4	84.5	65.0	72.0	88.2	71.4	68.0

Solution :

1. *First method using raw scores :*

(a) X and Y scores are totalled to give ΣX and ΣY respectively (Table 8.27).

(b) Each X score is squared and all such X^2 values are totalled to give ΣX^2 .

(c) The X score of each case is multiplied by the corresponding Y score and all such products are totalled to give ΣXY .

Table 8.27. Table for computing regression coefficient using raw scores.

Individual	X	Y	X^2	XY
1	75.2	3.0	5655.04	225.60
2	86.0	4.0	7396.00	344.00
3	78.0	3.3	6084.00	257.40
4	75.4	2.8	5685.16	211.12
5	84.5	3.5	7140.25	295.75
6	65.0	2.2	4225.00	143.00
7	72.0	2.5	5184.00	180.00
8	88.2	4.2	7779.24	370.44
9	71.4	2.3	5097.96	164.22
10	68.0	2.3	4624.00	156.40
Σ	763.7	30.1	58870.65	2347.93

(d) The data of Table 8.27 are used in computing b_{YX} .

$$b_{YX} = \frac{n \Sigma XY - \Sigma X \Sigma Y}{n \Sigma X^2 - (\Sigma X)^2} = \frac{10 \times 2347.93 - 763.7 \times 30.1}{10 \times 58870.65 - (763.7)^2} = 0.09.$$

2. *Alternative method using the sum of products :*

(a) The X and Y scores are totalled to give respectively ΣX and ΣY which are used in working out the respective means, \bar{X} and \bar{Y} (Table 8.28).

$$\bar{X} = \frac{\Sigma X}{n} = \frac{763.7}{10} = 76.4 ; \quad \bar{Y} = \frac{\Sigma Y}{n} = \frac{30.1}{10} = 3.01.$$

(b) Deviations of X scores from \bar{X} are worked out and squared, and these squared deviations are totalled to give the sum of squares, viz., $\Sigma(X - \bar{X})^2$. From Table 8.28,

$$\Sigma(X - \bar{X})^2 = 546.89.$$

Table 8.28. Table for computing regression coefficient using the sum of squares.

Individual	X	Y	$X - \bar{X}$	$(X - \bar{X})^2$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
1	75.2	3.0	- 1.2	1.44	- 0.01	0.012
2	86.0	4.0	+ 9.6	92.16	+ 0.99	9.504
3	78.0	3.3	+ 1.6	2.56	+ 0.29	0.464
4	75.4	2.8	- 1.0	1.00	- 0.21	0.210
5	84.5	3.5	+ 8.1	65.61	+ 0.49	3.969
6	65.0	2.2	- 11.4	129.96	- 0.81	9.234
7	72.0	2.5	- 4.4	19.36	- 0.51	2.244
8	88.2	4.2	+ 11.8	139.24	+ 1.19	14.042
9	71.4	2.3	- 5.0	25.00	- 0.71	3.550
10	68.0	2.3	- 8.4	70.56	- 0.71	5.964
Σ	763.7	30.1		546.89		49.193

(c) Deviation of each Y score from \bar{Y} is worked out and multiplied by the deviation of the corresponding X score from \bar{X} ; all such products are totalled give $\Sigma(X - \bar{X})(Y - \bar{Y})$. From Table 8.28,

$$\Sigma(X - \bar{X})(Y - \bar{Y}) = 49.193.$$

(d) The sum of squares and the sum of products are used in working out b_{YX} .

$$b_{YX} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2} = \frac{49.193}{546.89} = 0.09.$$

3. Computation of a_{YX} and the regression equation :

$$a_{YX} = \bar{Y} - b_{YX}\bar{X} = 3.01 - 0.09 \times 76.4 = - 3.87.$$

$$\hat{Y} = a_{YX} + b_{YX}X; \quad \text{or, } \hat{Y} = - 3.87 + 0.09 X.$$

8.13 MULTIPLE REGRESSION

Multiple regression, a method of *multivariate statistics*, predicts the most likely value of a variable (criterion or dependent variable) from the values of two or more other variables (predictors or independent variables). It can be computed only if the variables possess significant correlations with each other, It translates the relation between the criterion and a combination of two or more predictors into an expression showing the criterion as a

function of the predictors. Multiple regression may be *linear* or *nonlinear* according as the criterion is a linear or nonlinear function of the predictors. Multiple regressions belong to *models I, II and III* according as all the predictors are 'fixed' treatments, or all are classification variables, or some of them are treatment variables and others classification variables. The *Dubois-Dubois formula* for predicting the human body surface area from height and weight is a model II multiple regression equation.

Assumptions

It should be reasonable to assume that :

(a) all the variables involved are *continuous measurement variables* ;

(b) their scores have *unimodal and fairly symmetrical distributions* in the population ;

(c) the paired scores of each pair of variables in an individual are *independent* of all other such paired scores in the sample ;

(d) there is a *linear association* between the scores of each pair of variables ;

(e) the actual scores of the criterion in a large number of individuals possessing given predictor scores are *distributed normally*, independent of each other, around the criterion score predicted from those predictor scores and all such distributions obey *homoscedasticity*.

Multiple linear regression with three variables

Such a regression predicts the most likely value \hat{X}_1 of the criterion X_1 from the given values of two predictors X_2 and X_3 . The general regression equation for the straight line, showing X_1 as the linear function of X_2 and X_3 , is as follows :

$$\hat{X}_1 = a_{1.23} + b_{12.3} X_2 + b_{13.2} X_3$$

where $b_{12.3}$ and $b_{13.2}$ are the *partial regression coefficients*, and $a_{1.23}$ is the Y-intercept of the line. $b_{12.3}$ is the average rate of change of X_1 for each unit change of X_2 when the effect of the second predictor X_3 is *partialled out* ; $b_{12.3}$ thus gives the slope of the regression line of X_1 on X_2 when X_3 is held constant. The other coefficient $b_{13.2}$ is the average rate of change of X_1 for each unit change of X_3 when X_2 is *partialled out* ; $b_{13.2}$ thus gives the slope of the regression line of X_1 on X_3 when X_2 is kept constant. $b_{12.3}$ and $b_{13.2}$ are computed using the

SDs (s_1 , s_2 and s_3) of the respective variables and the *beta coefficients* (β_2 and β_3).

$$\beta_2 = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} ; \quad \beta_3 = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} ;$$

$$b_{12.3} = \beta_2 \times \frac{s_1}{s_2} ; \quad b_{13.2} = \beta_3 \times \frac{s_1}{s_3} ;$$

$$a_{1.23} = \bar{X}_1 - b_{12.3} \bar{X}_2 - b_{13.2} \bar{X}_3.$$

The value of $a_{1.23}$, $b_{12.3}$ and $b_{13.2}$ are then put in the preceding general formula for regression equation to give the required regression equation.

The SE of estimate ($s_{1.23}$) of X_1 on X_2 and X_3 is computed using the *multiple correlation coefficient* $R_{1.23}$.

$$R_{1.23} = \sqrt{\beta_2 r_{12} + \beta_3 r_{13}} ;$$

$$s_{1.23} = s_1 \sqrt{1 - R_{1.23}^2}.$$

Other 3-variable multiple regressions, e.g., regression of X_2 on X_1 and X_3 , are similarly worked out using specific beta coefficients.

Multiple linear regression with many variables

Where m is the total number of variables involved, the regression of X_1 on X_2, X_3, \dots, X_m is worked out as follows :

$$\hat{X}_1 = a_{1.23\dots m} + b_{12.34\dots m} X_2 + b_{13.24\dots m} X_3 + \dots + b_{1m.23\dots(m-1)} X_m ;$$

$$\text{or, } \hat{X}_1 = \bar{X}_1 + \beta_2 \frac{s_1}{s_2} (X_2 - \bar{X}_2) + \beta_3 \frac{s_1}{s_3} (X_3 - \bar{X}_3) + \dots + \beta_m \frac{s_1}{s_m} (X_m - \bar{X}_m).$$

The SE of estimate is worked out as follows, using the coefficient of multiple determination.

$$s_{1.23\dots m} = s_1 \sqrt{1 - R_{1.23\dots m}^2}.$$

Example 8.13.1.

Compute the multiple linear regression equation of arithmetic reasoning test scores (X_1) on the combination of numerical operation test scores (X_2) and mechanical knowledge test scores (X_3), using the following data in a sample of 103 students.

$$\begin{aligned}\bar{X}_1 &= 9.7, & s_1 &= 2.20; & \bar{X}_2 &= 44.0, & s_2 &= 8.80; & \bar{X}_3 &= 26.2, & s_3 &= 6.60; \\ r_{12} &= +0.46; & r_{13} &= +0.20; & r_{23} &= +0.17.\end{aligned}$$

Solution :

$$\beta_2 = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} = \frac{0.46 - 0.20 \times 0.17}{1 - 0.17^2} = 0.44.$$

$$\beta_3 = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} = \frac{0.20 - 0.46 \times 0.17}{1 - 0.17^2} = 0.13.$$

$$b_{12.3} = \beta_2 \frac{s_1}{s_2} = 0.44 \times \frac{2.20}{8.80} = 0.11.$$

$$b_{13.2} = \beta_3 \frac{s_1}{s_3} = 0.13 \times \frac{2.20}{6.60} = 0.043 \approx 0.04.$$

$$a_{1.23} = \bar{X}_1 - b_{12.3}\bar{X}_2 - b_{13.2}\bar{X}_3 = 9.7 - 0.11 \times 44.0 - 0.043 \times 26.2 = 3.73.$$

The partial regression coefficients and $a_{1.23}$ are then used to compute the regression equation.

$$\hat{X}_1 = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3;$$

$$\text{or, } \hat{X}_1 = 3.73 + 0.11X_2 + 0.04X_3.$$

$$R_{1.23} = \sqrt{\beta_2 r_{12} + \beta_3 r_{13}} = \sqrt{0.44 \times 0.46 + 0.13 \times 0.20} = 0.478.$$

$$\therefore s_{1.23} = s_1 \sqrt{1 - R_{1.23}^2} = 2.20 \sqrt{1 - (0.478)^2} = 1.932.$$

Example 8.13.2.

In a sample of 40 rats, the means and standard deviations of thyroid calcitonin content (X_1), thyroxine secretion rate (X_2) and serum calcium concentration (X_3) were found to be as follows :

$$\bar{X}_1 = 395 \text{ mU/gland}; \quad s_1 = 72.16 \text{ mU};$$

$$\bar{X}_2 = 1.10 \text{ } \mu\text{g}/100 \text{ g bodyweight}; \quad s_2 = 0.32 \text{ } \mu\text{g};$$

$$\bar{X}_3 = 8.2 \text{ mg/dL}; \quad s_3 = 2.05 \text{ mg/dL}.$$

The product-moment r values between the respective variables of each pair were worked out as follows:

$$r_{12} = -0.38; \quad r_{13} = +0.65; \quad r_{23} = -0.12.$$

Work out the regression equation of thyroidal calcitonin on the combination of thyroxine secretion rate and serum calcium.

Solution :

(a) The beta coefficients are computed from the product-moment r values.

$$\beta_2 = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} = \frac{-0.38 - 0.65 \times (-0.12)}{1 - (-0.12)^2} = -0.31 ;$$

$$\beta_3 = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} = \frac{0.65 - (-0.38) \times (-0.12)}{1 - (-0.12)^2} = 0.61.$$

(b) The partial regression coefficients are computed using the beta coefficients and standard deviations.

$$b_{12.3} = \beta_2 \frac{s_1}{s_2} = -0.31 \times \frac{72.16}{0.32} = -69.91 ; \quad b_{13.2} = \beta_3 \frac{s_1}{s_3} = 0.61 \times \frac{72.16}{2.05} = 21.47.$$

(c) The partial regression coefficients and the means are used in computing $a_{1.23}$ as follows :

$$a_{1.23} = \bar{X}_1 - b_{12.3}\bar{X}_2 - b_{13.2}\bar{X}_3 = 395 - (-69.91) \times 1.10 - 21.47 \times 8.2 = 295.85.$$

(d) The multiple regression equation of X_1 on X_2 and X_3 may be written as follows :

$$\hat{X}_1 = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3 ; \quad \therefore \hat{X}_1 = 295.85 - 69.91X_2 + 21.47X_3.$$

(e) The SE of estimate ($s_{1.23}$) is computed using the multiple correlation coefficient ($R_{1.23}$).

$$R_{1.23} = \sqrt{\beta_2 r_{12} + \beta_3 r_{13}} = \sqrt{-0.31 \times (-0.38) + 0.61 \times 0.65} = 0.717 ;$$

$$\therefore s_{1.23} = s_1 \sqrt{1 - R_{1.23}^2} = 72.16 \sqrt{1 - (0.717)^2} = 50.30.$$

GLOSSARY

biserial r : a modified form of product-moment r for linear correlation between the scores of a continuous variable and the classes of an apparently dichotomous or artificially dichotomized variable.

coefficient of determination : measure of that proportion of the total variance of a variable, which is associated with the variance of a correlated variable.

coefficient of nondetermination : measure of that proportion of the total variance of a variable, which is not associated with the variance of a correlated variable.

contingency coefficient : a nonparametric correlation coefficient for correlating two discontinuous or nominal variables divided into more than two classes.

correlation : statistics for establishing and estimating both the degree and the direction of association between two or more variables.

correlation, linear : correlation between variables having an association conforming to a straight line and giving a linear scattergram.

correlation, multiple : correlation between a given variable and the weighted sum of two or more other variables.

correlation, negative : correlation where the changes in the magnitude of one variable are associated with the reverse changes of the other variable.

correlation, partial : correlation between two variables, eliminating the effects of one or more other variables correlated with them.

correlation, positive : correlation where the changes in the magnitude of one variable are associated with the changes, in the same direction, of the other variable.

correlation, simple : correlation between only two variables.

correlation coefficient : a statistic for establishing and estimating both the magnitude and the direction (algebraic sign) of the association between two or more variables.

criterion : the dependent variable whose most likely value is predicted by regression in an individual or case, using the known value(s) of one or more predictors (independent variables) in the latter.

homoscedasticity : homogeneity of the variances of samples drawn from the same (or similar) population(s) so that such variances may serve as estimates of the same population variance and their differences can be explained away by their sampling errors.

Kendall's tau : a nonparametric correlation coefficient for simple linear correlation between the ranks of the individuals or cases with respect to the two corresponding variables.

phi coefficient : a nonparametric correlation coefficient for correlating the classes of two genuinely dichotomous variables.

point biserial r : a modified form of product-moment r for linear correlation between the scores of a continuous variable and the classes of a genuinely dichotomous variable.

predictor : an independent variable whose known value in an individual or case is used for predicting in the latter the most likely value of a criterion (dependent variable).

product-moment r : a correlation coefficient for simple linear correlation between the scores of two continuous variables with normal or near-normal distributions in the population.

regression : statistical prediction of the most likely value of a criterion in an individual or case, using the known value(s) of one or more predictors in the latter.

regression, linear : regression where the criterion has linear correlation(s) with the predictor(s).

regression, model I : regression using "fixed" treatment variable(s) as predictor(s).

regression, model II : regression using classification variable(s) as predictor(s).

regression, model III : multiple regression using a combination of both treatment and classification variables as predictors.

regression, multiple : regression using two or more predictors.

regression, simple : regression using a single predictor.

regression coefficient : a prediction statistic which estimates and expresses the average rate of change in the value of a criterion for unit changes in the value of a predictor.

Spearman's rho : a nonparametric correlation coefficient for simple linear correlation between the ranks of the individuals or cases with respect to the two corresponding variables.

standard error of estimate : a statistic estimating the differences of the predicted value of a criterion from the actual values of the latter in individuals or cases possessing a given value of the predictor.

sum of products : sum of the products of deviations of the scores of two variables, in each individual case of the sample, from their respective means.

sum of squares : sum of the squared deviations of the scores of a variable from a central value such as the mean.

tetrachoric r : a correlation coefficient for correlating the classes of two apparently dichotomous or artificially dichotomized variables.

9. NONPARAMETRIC STATISTICS

Nonparametric statistics require few assumptions, no estimate of parameter in their computation, and no normal distribution of the variables in the population.

9.1 NONPARAMETRIC STATISTICS

Parametric statistics serve as estimates of the corresponding parameters. Their computations require the use of *precomputed statistics* as estimates of parameters. Moreover, they are interpreted with reference to *specific population distributions* of the variables such as the normal and *t* distributions. Thus, they cannot be used for too small samples, nominal, ordinal and discrete variables, and non-normal distributions.

Nonparametric (distribution-free) statistics (i) require very few assumptions, (ii) do not require normal distributions of the variables in the population, (iii) do not use any precomputed statistic as an estimate of parameter in the computation, (iv) can be used for even very small samples and (v) are computed by much simpler methods ; but (vi) their powers are lower than their parametric counterparts so long as the assumptions for the latter are fulfilled.

Spearman's and Kendall's rank correlations as well as phi and contingency coefficients, discussed in chapter 8, are nonparametric statistics. Some other nonparametric statistics are described here.

9.2 CHI SQUARE TESTS

Nonparametric chi square (χ^2) test explores the significance of deviation of an experimentally observed frequency distribution from a proposed frequency distribution and,

therefore, constitutes an *analysis of frequencies*. It (i) requires no assumption for normality of the population distribution of variable(s), (ii) uses no precomputed statistic as an estimate of parameter in its computation, (iii) is applicable to very small samples, and (iv) can be used also for discrete, nominal or ordinal variables.

χ^2 is the sum of the ratios of (i) squared deviations of observed frequencies (f_o) from the frequencies (f_e) expected from a given distribution, and (ii) the respective f_e values.

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

χ^2 has positively skewed probability distributions (Fig. 9.1) depending on its *df*. The area under the χ^2 distribution curve for the given *df* is taken as 1.00. Each χ^2 distribution is asymptotic to the abscissa in its right tail — its right tail reaches the abscissa at $+\infty$ only. χ^2 cannot have negative values. When $df \approx \infty$, the distribution is very little skewed ; indeed, it approaches almost a bilateral symmetry at

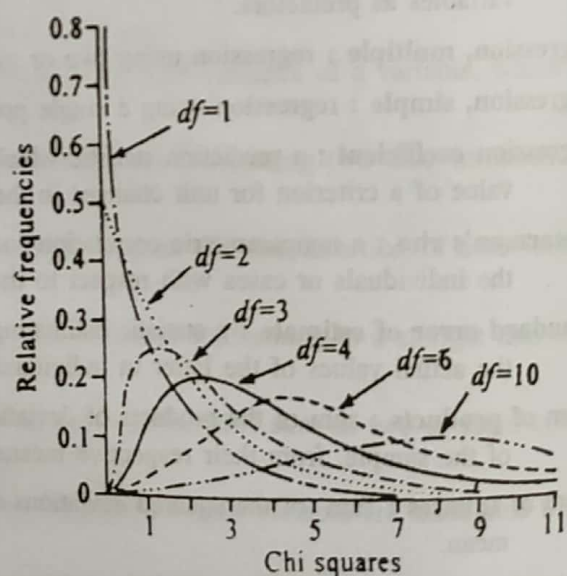


Fig. 9.1. Some probability distributions of chi squares.

$df > 30$. With the fall in df , positive skewness increases progressively. At $df = 1$, the distribution is L-shaped or reverse J-shaped, asymptotic to both X and Y axes. Thus, the probability of getting a given χ^2 by mere chance due to random sampling equals the fractional area beyond the ordinate at the given χ^2 in the right tail of the distribution of χ^2 values with the specified df .

1. Chi square test of independence

An *association* may exist between two variables if the change, in value or quality, of one variable is accompanied by a similar or opposite change of the other in the same individual. Absence of association is called the *independence* of two variables. A χ^2 test of independence (or, of association) explores whether or not two variables are significantly independent of each other; in other words, it explores the existence of any significant association between the variables. But it can (i) neither measure the magnitude and the direction of the association, (ii) nor predict a cause-and-effect relationship between the variables.

The data consist of frequencies of paired observations of two variables, often discrete or nominal, distributed in different combinations of their classes. χ^2 is computed in the following way.

(a) The frequencies of paired observations are first arranged in a *contingency table* showing the association between two variables in their combined distribution. A contingency table is a two-way classification table with the classes of one variable arranged in r number of rows and those of the other in c number of columns (Table 9.1). Each cell of the table represents a specific combination of two classes, one of each variable. According to the number of rows and columns, i.e., the number of classes of the variables, the table is an $r \times c$ -fold table. For example, it is a 3×2 -fold

contingency table where one variable has three classes as against two classes of the other. The degrees of freedom (df) of the computed χ^2 are given by: $df = (r - 1)(c - 1)$.

The observed frequencies (f_o) of cases, belonging to the specific combinations of classes of the variables, are entered in the corresponding cells of the table. The total frequency of a row or column, f_r or f_c respectively, is entered respectively in the marginal column and row; these total frequencies are called the respective *marginal frequencies*. Where n is the total frequency or sample size,

$$\Sigma f_r = \Sigma f_c = n.$$

(b) The expected frequency (f_e) of each cell is computed on the basis of the H_0 and entered against the corresponding f_o (Table 9.2) — the H_0 for a test of independence proposes that the variables are independent of each other, that there is no significant difference between the distributions of f_o and f_e scores, and that any ($f_o - f_e$) difference has resulted from mere chances of random sampling and would not have occurred if the entire population was studied instead of a random sample. However, to keep the sample size (n) constant, the f_e values are directly computed, according to the following formula in terms of the H_0 , for only as many randomly chosen cells as the df of the χ^2 .

$$f_e = \frac{f_r f_c}{n},$$

where f_r and f_c are marginal total frequencies of respectively the row and the column to which the given cell belongs. The f_e of each of the remaining cells is computed by subtracting the already computed f_e values from the marginal totals.

(c) The difference ($f_o - f_e$) is worked out for each cell and entered in the table.

(d) Each such difference is squared to give the corresponding $(f_o - f_e)^2$.

(e) Each of the latter is divided by th

corresponding f_e to get the ratio $(f_o - f_e)^2 / f_e$. The sum of all these ratios are used in working out χ^2 .

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Yates' correction :

If any computed f_e is less than 5 and the χ^2 has the *df* of 1 only, Yates' correction has to be applied. Yates' correction brings each $(f_o - f_e)$ closer to zero by 0.5 — this means the subtraction of 0.5 from each positive $(f_o - f_e)$ and the addition of 0.5 to each negative $(f_o - f_e)$. The corrected $(f_o - f_e)$ values are used for computing χ^2 . In other words,

$$\chi^2 = \sum \frac{(|f_o - f_e| - 0.5)^2}{f_e}$$

where the bars on two sides of $f_o - f_e$ indicate that all values of $(f_o - f_e)$ are taken as positive, ignoring their algebraic signs.

Alternative formula for fourfold contingency tables :

For the 2×2 -fold contingency tables, the following alternative method may be used avoiding the computation of $(f_o - f_e)$ values.

(a) In the 2×2 -fold contingency table, the classes of the two variables are arranged exactly like that for working out the phi coefficient (page 181). Thus, the cell A at the top right corner and the cell D at the bottom left corner contain *concordant cases* belonging either to the high value classes of both variables (cell A) or to the low value classes of both (cell D); the cell B at the top left corner and the cell C at the bottom right corner contain *discordant cases* belonging to the high value class of one variable and the low value class of the other, or vice versa (Table 9.3).

(b) The f_o values for different class combinations are then entered in the respective cells; marginal totals are also worked out for each row or column and entered as the f_r or f_c

(page 181 and Table 9.3). Evidently, the f_r for the cells A and B amounts to the total of their cell frequencies, viz., $A + B$, and that for the cells C and D is given by the sum of their frequencies, viz., $C + D$; similarly, the f_c for B and D amounts to $B + D$, while the f_c for cells A and C is the sum of their frequencies, viz., $A + C$.

(c) χ^2 is then computed using the products of pairs of cell frequencies (AD and BC) and the marginal totals.

$$\chi^2 = \frac{n(AD - BC)^2}{(A + B)(A + C)(B + D)(C + D)}$$

$$df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1.$$

If it seems that any f_e may be lower than 5, this is checked by computing the lowest f_e possible for any cell, using the smaller of the f_r values and the smaller of the f_c values.

$$\text{lowest } f_e = \frac{(\text{smaller } f_r) \times (\text{smaller } f_c)}{n}$$

When any f_e is found to be less than 5, Yates' correction is applied by changing the computational formula :

$$\chi^2 = \frac{n(|AD - BC| - n/2)^2}{(A + B)(A + C)(B + D)(C + D)}$$

where the bars on the two sides of $AD - BC$ indicate that $(AD - BC)$ is taken as positive irrespective of its algebraic sign.

Significance of computed χ^2 :

The H_0 contends that the computed χ^2 is not significant and has resulted from mere chances of random sampling. To test this H_0 by a two-tail test, the computed χ^2 is compared to the critical χ^2 with the computed *df* and for the chosen level of significance (α). Table C of Appendix lists critical χ^2 values according to their *df* and α for two-tail tests. If the computed χ^2 exceeds or equals the critical $\chi^2_{\alpha(df)}$, the probability P of obtaining the computed χ^2 by chance is respectively lower than and equal to α and may be considered too

low ($P \leq \alpha$). The H_0 is then rejected and the computed χ^2 is considered significant — the variables are then considered to have a significant association with each other. But if the computed χ^2 is lower than the critical $\chi^2_{\alpha(df)}$, P exceeds α and is too high to justify the rejection of H_0 ($P > \alpha$); the computed χ^2 is then not significant — the variables have no significant association with each other.

For a one-tail χ^2 test, α for a given critical χ^2 is half that of an identical critical χ^2 for two-tail tests. In other words, the one-tail critical χ^2 for a given significance level is identical with the two-tail critical χ^2 for double that significance level. Thus, the one-tail critical χ^2 ($df = 4$) for 0.05 level amounts to 7.78 which is identical with the two-tail critical χ^2 ($df = 4$) for 0.10 level.

Example 9.2.1.

Using the data presented in Table 9.1 about frequencies of gene crossovers between homologous chromosomes in *Drosophila*, born of mothers of different ages, find if there is a significant association between the frequency of crossovers and the mother's age.

Table 9.1. 4×2 -fold contingency table showing crossover data.

Mother's age (days)	Crossovers (f_o)	Noncrossovers (f_o)	Total (f_r)
5	247	291	538
10	152	284	436
20	174	393	567
30	140	495	635
Total (f_c)	713	1463	2176 (n)

Solution :

A two-tail χ^2 test is applied for testing the H_0 which proposes that there is no significant association between crossover frequency and mother's age.

(a) The f_o values are entered in a 4×2 -fold contingency table (Table 9.2). The marginal totals (f_r and f_c) are worked out and entered.

$$df = (r - 1)(c - 1) = (4 - 1)(2 - 1) = 3.$$

(b) The f_e is calculated for each of as many cells as the df (viz., 3) and entered in Table 9.2 against the respective f_o values. For example, for the f_o of 247 of the cell for crossover vs 5 days' age,

$$f_e = \frac{f_r f_c}{n} = \frac{538 \times 713}{2176} = 176.3.$$

The f_e values corresponding to the f_o values of 174 and 284 are similarly computed to be 185.8 and 293.1, respectively.

(c) The remaining f_e values are obtained by subtracting the sum of the already computed f_e values from either f_r or f_c .

(d) The $(f_o - f_e)$ value is computed for each f_o and entered in Table 9.2. For example, corresponding to the f_o of 247,

$$f_o - f_e = 247 - 176.3 = 70.7.$$

(e) χ^2 is then computed using each $(f_o - f_e)^2$ and the corresponding f_e .

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(70.7)^2}{176.3} + \frac{(9.1)^2}{142.9} + \frac{(-11.8)^2}{185.8} + \frac{(-68.0)^2}{208.0} + \frac{(-70.7)^2}{361.7} + \frac{(-9.1)^2}{293.1} + \frac{(11.8)^2}{381.2} + \frac{(68.0)^2}{427.0} = 77.21.$$

(f) Critical χ^2 values ($df = 3$) for different levels of α are quoted from Table C of Appendix.

$$\chi^2_{.05(3)} = 7.82 ; \quad \chi^2_{.01(3)} = 11.34 ; \quad \chi^2_{.001(3)} = 16.27.$$

The computed χ^2 of 77.21 far exceeds even the critical χ^2 for 0.001 level. So, the probability P of getting the computed χ^2 by chance due to random sampling is far lower than 0.001. Hence, the probability is too low for the H_0 to be correct. So, the H_0 is rejected and the computed χ^2 is considered significant — the frequency of crossovers, therefore, has a *significant association* with the mother's age ($P < 0.001$).

Table 9.2. 4×2 -fold contingency table for computing χ^2 of crossover data.

Mother's age	Crossovers			Noncrossovers			f_r
	f_o	f_e	$f_o - f_e$	f_o	f_e	$f_o - f_e$	
5	247	176.3	+ 70.7	291	361.7	- 70.7	538
10	152	142.9	+ 9.1	284	293.1	- 9.1	436
20	174	185.8	- 11.8	393	381.2	+ 11.8	567
30	140	208.0	- 68.0	495	427.0	+ 68.0	635
f_c	713	713.0		1463	1463.0		2176 (n)

Example 9.2.2.

Responses of 48 normal humans and 57 neurotics to an item of a neurotic questionnaire are tabulated in Table 9.3. Find whether or not the test item differentiates significantly the normal humans from neurotics.

Solution :

A two-tail χ^2 test is applied to determine whether or not the test item has a significant association with neurotic condition.

Table 9.3. Fourfold (2×2) contingency table for neurotic questionnaire data.

Subjects	Negative response (f_o)	Positive response (f_o)	Total (f_r)
Normal	28 (B)	20 (A)	48 (A + B)
Neurotic	20 (D)	37 (C)	57 (C + D)
Total (f_c)	48 (B + D)	57 (A + C)	105 (n)

$$\chi^2 = \frac{n(AD - BC)^2}{(A + B)(A + C)(B + D)(C + D)} = \frac{105(20 \times 20 - 28 \times 37)^2}{48 \times 57 \times 48 \times 57} = 5.67.$$

$$df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1.$$

The computed χ^2 is next compared with critical χ^2 values ($df = 1$) for different levels of significance (Table C of Appendix) :

$$\chi^2_{.05(1)} = 3.84 ; \quad \chi^2_{.02(1)} = 5.41 ; \quad \chi^2_{.01(1)} = 6.64.$$

Because the computed χ^2 of 5.67 is higher than the critical $\chi^2_{.02}$, the probability P of getting the computed χ^2 by chance of random sampling is lower than 0.02. So, the computed χ^2 is significant at 0.02 level. Thus, there is a *significant association* between the test item and the neurotic condition ($P < 0.02$). Hence, the test item *significantly differentiates* the normal from the neurotic.

Example 9.2.3.

Out of 15 hypertensive human subjects exposed to cold during a cold pressor test, 4 reacted with 10-20 mm Hg rise in the diastolic pressure (normoreactors), 9 showed more than 20 mm Hg rise (hyperreactors) and 2 reacted with less than 10 mm Hg rise (hyporeactors). Out of 25 nonhypertensive subjects similarly treated, 12 were normoreactors, 5 were hyperreactors and 8 were hyporeactors. Is there any significant association between the cold pressor reaction and hypertension ?

Solution :

A two-tail χ^2 test is applied for testing the H_0 which contends that there is no significant association between hypertension and cold pressor reaction.

(a) The f_o values are entered in a 3×2 -fold contingency table and the marginal totals (f_r and f_c) are computed and entered therein (Table 9.4).

$$df = (r - 1)(c - 1) = (3 - 1)(2 - 1) = 2.$$

Table 9.4. 3×2 -fold contingency table for cold pressor reaction data.

Cold pressor reaction	Hypertensive			Nonhypertensive			f_r
	f_o	f_e	$f_o - f_e$	f_o	f_e	$f_o - f_e$	
Normoreactor	4	6.00	- 2.00	12	10.00	+ 2.00	16
Hyperreactor	9	5.25	+ 3.75	5	8.75	- 3.75	14
Hyporeactor	2	3.75	- 1.75	8	6.25	+ 1.75	10
Total (f_c)	15	15.00		25	25.00		40 (n)

(b) The f_e values are calculated for as many cells as the df (viz., 2) and entered in the table. For example, for the f_o of 9 in the cell for hypertensive vs hyperreactor.

$$f_e = \frac{f_r f_c}{n} = \frac{14 \times 15}{40} = 5.25.$$

In a similar way, the f_e corresponding to the f_o of 2 in the cell for hypertensive vs hyporeactor, is computed to be 3.75.

(c) The f_e values of the remaining cells are then obtained by subtracting the sums of the already computed f_e values from either f_r or f_c .

(d) The ($f_o - f_e$) values are computed for each f_o and entered in Table 9.4. For example, for the f_o of 9,

$$f_o - f_e = 9 - 5.25 = + 3.75.$$

(e) One f_e value, viz., 3.75, is lower than 5 ; but $df > 1$. So, Yates' correction is not applied.

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(-2.00)^2}{6.00} + \frac{(3.75)^2}{5.25} + \frac{(-1.75)^2}{3.75} + \frac{(2.00)^2}{10.00} + \frac{(-3.75)^2}{8.75} + \frac{(1.75)^2}{6.25}$$

$$= 6.66.$$

(f) The critical χ^2 values ($df = 2$) for different levels of α are quoted from Table C of Appendix.

$$\chi^2_{.05(2)} = 5.99 ; \quad \chi^2_{.02(2)} = 7.82 ; \quad \chi^2_{.01(2)} = 9.21.$$

The computed χ^2 of 6.66 is higher than the critical χ^2 for 0.05 level. So, the probability P of getting the computed χ^2 by chance due to random sampling is lower than 0.05. The computed χ^2 is, therefore, considered significant. Thus, hypertension has a *significant association* with the cold pressor reaction ($P < 0.05$).

Example 9.2.4.

Out of 15 diabetic subjects, 8 were found to be suffering from hypercholesterolemia while the rest had normal serum cholesterol. Out of 10 nondiabetics, 2 had high serum cholesterol while the rest had normal serum cholesterol level. Is there any significant association between hypercholesterolemia and diabetes ?

Solution :

1. *First method avoiding $(f_o - f_e)$:*

(a) The f_o values are entered in a 2×2 -fold contingency table and the marginal totals (f_r and f_c) are also computed and entered therein (Table 9.5).

$$df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1.$$

Table 9.5. Fourfold (2×2) contingency table for diabetes data, avoiding $(f_o - f_e)$.

	Nonhypercholesterolemic (f_o)	Hypercholesterolemic (f_o)	Total (f_r)
Diabetic	7 (B)	8 (A)	15 (A + B)
Nondiabetic	8 (D)	2 (C)	10 (C + D)
Total (f_c)	15 (B + D)	10 (A + C)	25 (n)

(b) To test whether or not any cell has an f_e lower than 5, the f_e is computed for the cell corresponding to the lower values of both f_r and f_c . Thus for the cell C in the present case,

$$f_e = \frac{f_r f_c}{n} = \frac{10 \times 10}{25} = 4.$$

(c) As $df = 1$, and one of the cells has an f_e lower than 5, Yates' correction is applied in computing the χ^2 .

$$\chi^2 = \frac{n(|AD - BC| - n/2)^2}{(A+B)(A+C)(B+D)(C+D)} = \frac{25(|8 \times 8 - 7 \times 2| - 25/2)^2}{15 \times 10 \times 15 \times 10} = 1.56.$$

2. *Alternative method using $(f_o - f_e)$:*

(a) The f_o values are entered in the relevant columns of a 2×2 -fold contingency table and the marginal

totals (f_r and f_c) are also worked out and entered therein (Table 9.6).

$$df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1.$$

(b) As $df = 1$, the f_e for any one of the cells is worked out using the relevant f_r and f_c values and entered in the table. For example, for the cell for nondiabetic vs hypercholesterolemic,

$$f_e = \frac{f_r f_c}{n} = \frac{10 \times 10}{25} = 4.$$

(c) The f_e of each of the remaining cells is worked out by subtracting the already computed f_e values from either f_r or f_c .

(d) As $df = 1$, and one f_e value (viz., 4) is less than 5, Yates' correction is applied to all the $(f_o - f_e)$. To do this, 0.5 is added to each negative $(f_o - f_e)$ and subtracted from each positive $(f_o - f_e)$ to give the respective "corrected $(f_o - f_e)$ " values (Table 9.6).

Table 9.6. Fourfold contingency table for diabetes data, using $(f_o - f_e)$.

	Nonhypercholesterolemic				Hypercholesterolemic				Total (f_r)
	f_o	f_e	$f_o - f_e$	Corrected $(f_o - f_e)$	f_o	f_e	$f_o - f_e$	Corrected $(f_o - f_e)$	
Diabetic	7	9	- 2	- 1.5	8	6	+ 2	+ 1.5	15
Nondiabetic	8	6	+ 2	+ 1.5	2	4	- 2	- 1.5	10
Total (f_c)	15	15			10	10			25 (n)

(e) The squared "corrected $(f_o - f_e)$ " values are used in working out χ^2 .

$$\chi^2 = \sum \left[\frac{\text{corrected } (f_o - f_e)^2}{f_e} \right] = \frac{(-1.5)^2}{9} + \frac{(1.5)^2}{6} + \frac{(1.5)^2}{6} + \frac{(-1.5)^2}{4} = 1.56.$$

Interpretation :

To compare with the χ^2 worked out by any of the preceding methods, the critical χ^2 values ($df = 1$) for different levels of α are quoted from Table C of Appendix.

$$\chi^2_{0.05(1)} = 3.84 ; \quad \chi^2_{0.02(1)} = 5.41 ; \quad \chi^2_{0.01(1)} = 6.64.$$

As the computed χ^2 is lower than even the critical $\chi^2_{0.05}$, the probability is too high that the computed χ^2 has been obtained by chance due to random sampling ($P > 0.05$). Hence, the H_0 cannot be rejected. So, the computed χ^2 is not significant — there is no significant association between diabetes and hypercholesterolemia.

2. Chi square test for goodness of fit

This test is used to explore how far a distribution of observed frequencies (f_o) fits with a theoretical distribution such as the normal distribution, a binomial distribution, a Mendelian phenotype distribution, and a distribution proposed by the hypothesis of equal probability. The f_e values are computed

here on the basis of the proposed theoretical distribution. The χ^2 computed from the $(f_o - f_e)$ values proves significant if it equals or exceeds the critical χ^2 for the chosen level of significance ; in such a case, the f_o distribution differs significantly from the proposed distribution. A computed χ^2 lower than the critical χ^2 indicates that the f_o distribution fits

with the proposed distribution and does not differ significantly from the latter.

The classical formula, based on $(f_o - f_e)$ values, is used in computing the χ^2 .

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

The alternative formula, avoiding the use of f_e values, cannot be applied as no contingency table is involved. *Yates' correction* (page 204) has to be applied if the f_e of any class falls below 5 and the df amounts to 1 only.

Degrees of freedom :

The df of the computed χ^2 depends in this test on the nature of the proposed theoretical

distribution and is limited to the number of classes in the distribution whose frequencies are free to vary. Distributions like the *Mendelian 9 : 3 : 3 : 1 phenotype distribution* and the equal probability hypothesis distribution do not involve any parameter ; so, only one df is lost in keeping the sample size (n) constant — for such distributions, $df = k - 1$, where k is the number of classes in the distribution. For distributions involving estimates of one or more parameters, the df is further lowered by the number of parameters involved. For example, for a *normal distribution*, df amounts to $k - 3$, because three degrees of freedom are lost in keeping n , μ and σ constant ; for *binomial* and *Poisson distributions*, df amounts to $k - 2$, because n and μ are to be kept constant.

Example 9.2.5.

Find whether or not the observed distribution of serum iron concentration, presented in *Example 6.4.1*, differs significantly from the normal distribution.

Solution :

(a) The expected frequencies of a *best-fitting normal distribution* are first computed (§ 6.4). In the present case, this has already been done stepwise in *Example 6.4.1* and presented in Table 6.4 (page 84). The observed and expected frequencies of that table are entered in Table 9.7 for further computations.

Table 9.7. Table for the goodness of fit of serum iron with best-fitting normal distribution.

Classes	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
100-109	6	3.5	+ 2.5	6.25	1.786
110-119	11	8.4	+ 2.6	6.76	0.805
120-129	10	14.7	- 4.7	22.09	1.503
130-139	17	18.4	- 1.4	1.96	0.107
140-149	16	16.4	- 0.4	0.16	0.010
150-159	13	10.5	+ 2.5	6.25	0.595
160-169	7	4.8	+ 2.2	4.84	1.008
Total	80				5.814

(b) For each class, the values of $(f_o - f_e)$, $(f_o - f_e)^2$ and $(f_o - f_e)^2/f_e$ are worked out and entered in the respective columns of Table 9.7. For example, for the class 130-139,

$$f_o - f_e = 17 - 18.4 = -1.4 ; \quad (f_o - f_e)^2 = (-1.4)^2 = 1.96 ; \quad \frac{(f_o - f_e)^2}{f_e} = \frac{1.96}{18.4} = 0.107.$$

As $df > 1$, Yates' correction is not used inspite of some f_e values being lower than 5.

$$df = k - 3 = 7 - 3 = 4,$$

where k is the number of classes and three of them lose their freedom for variation to keep n , μ and σ constant for the normal distribution.

(c) From Table 9.7,

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 5.81.$$

(d) the critical χ^2 values ($df = 4$) for different levels of α are quoted from Table C of Appendix.

$$\chi_{0.05(4)}^2 = 9.49 ; \quad \chi_{0.02(4)}^2 = 11.67 ; \quad \chi_{0.01(4)}^2 = 13.28 ; \quad \chi_{0.001(4)}^2 = 18.46.$$

As the computed χ^2 is lower than the critical χ^2 for 0.05 level, the computed χ^2 is not significant ($P > 0.05$). So, there is *no significant difference* between the observed distribution and the best-fitting normal distribution. Hence, the observed distribution has a *significant goodness of fit* with the normal distribution.

Example 9.2.6.

Crossing a grey-bodied scarlet-eyed *Drosophila* with a black-bodied red-eyed one produced all grey-bodied red-eyed flies in the F_1 generation. On crossing the F_1 flies, the F_2 generation gave following phenotypes : grey-bodied red-eyed = 362 ; black-bodied red-eyed = 128 ; grey-bodied scarlet-eyed = 122 ; black-bodied scarlet-eyed = 44. Do the data have a goodness of fit with the Mendelian 9 : 3 : 3 : 1 distribution ?

Solution :

(a) The f_o of each phenotype is entered in Table 9.8 and the corresponding f_e is computed on the basis of the 9 : 3 : 3 : 1 distribution so that the f_e values of phenotypes conform to the respective proportions in that distribution. For example, for the grey red phenotype,

$$n = 362 + 128 + 122 + 44 = 656 ; \quad f_e = \frac{9}{9+3+3+1} \times n = \frac{9}{16} \times 656 = 369.$$

(b) For each phenotype, the values of $(f_o - f_e)$, $(f_o - f_e)^2$ and $(f_o - f_e)^2/f_e$ are computed and entered in the respective columns of Table 9.8. For example, for the black scarlet phenotype,

$$f_o - f_e = 44 - 41 = +3 ; \quad (f_o - f_e)^2 = 3^2 = 9 ; \quad \frac{(f_o - f_e)^2}{f_e} = \frac{9}{41} = 0.2195.$$

(c) From Table 9.8,

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 0.5637 ; \quad df = k - 1 = 4 - 1 = 3,$$

where k is the number of phenotypes, and one df is lost in keeping n constant for the Mendelian distribution.

Table 9.8. Computation of χ^2 for goodness of fit with Mendelian 9:3:3:1 distribution.

Phenotypes	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
Grey red	362	369	- 7	49	0.1328
Black red	128	123	+ 5	25	0.2033
Grey scarlet	122	123	- 1	1	0.0081
Black scarlet	44	41	+ 3	9	0.2195
Total	656 (n)	656			0.5637 (χ^2)

(d) Critical χ^2 values ($df = 3$) for different levels of significance are quoted from Table C of Appendix.

$$\chi^2_{0.05(3)} = 7.82 ; \quad \chi^2_{0.02(3)} = 9.84 ; \quad \chi^2_{0.01(3)} = 11.34 ; \quad \chi^2_{0.001(3)} = 16.27.$$

As the computed χ^2 of 0.5637 is far lower than the critical χ^2 for 0.05 level, the computed χ^2 is not significant ($P > 0.05$). So, there is *no significant difference* between the observed distribution and the 9:3:3:1 distribution — the observed frequency distribution of phenotypes has a *significant goodness of fit* with the proposed Mendelian distribution.

Example 9.2.7.

Crossing of a purple-eyed straight-winged *Drosophila* with a red-eyed curved-winged *Drosophila* produced dihybrid red-eyed straight-winged females in the F_1 generation. On crossing such F_1 females with double-recessive purple-eyed curved-winged males, the F_2 generation showed the following phenotype distribution : red straight = 339 ; purple straight = 612 ; red curved = 725 ; purple curved = 384. Red eyes and straight wings are dominant factors (A and B respectively) while purple eyes and curved wings are recessive factors (a and b respectively).

Find (i) whether or not the F_2 generation obeys the Mendelian law of independent assortment, and (ii) whether or not there is an indication of linkage between the two relevant genes.

Solution :

1. χ^2 test for independent assortment :

It is a χ^2 test for goodness of fit. The H_0 proposes that the ratio of frequencies in the F_2 generation would not have differed from 9:3:3:1 if the entire population were studied instead of a sample. This would mean that a two-factor independent assortment has occurred. To test this H_0 , χ^2 is computed as follows.

(a) The f_o of each phenotype is entered in Table 9.9.

Table 9.9. Computation of χ^2 for independent assortment.

Phenotypes	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
Red straight (AB)	339	1158.75	- 819.75	671990.06	579.93
Purple straight (aB)	612	386.25	+ 225.75	50963.06	131.94
Red curved (Ab)	725	386.25	+ 338.75	114751.56	297.09
Purple curved (ab)	384	128.75	+ 255.25	65152.56	506.04
Total	2060 (n)	2060.00			1515.00 (χ^2)

(b) In terms of the proposed 9:3:3:1 distribution, the proportions of the total frequency (n) in the respective phenotypes should be as follows : AB = 9/16, aB = 3/16, Ab = 3/16, ab = 1/16. These proportions are multiplied by the sample size n to give the respective f_e values. For example, for the AB phenotype,

$$n = 339 + 612 + 725 + 384 = 2060 ; \quad f_e = \frac{9}{16} \times 2060 = 1158.75.$$

(c) For each phenotype, $(f_o - f_e)$, $(f_o - f_e)^2$ and $(f_o - f_e)^2/f_e$ are computed and entered in the table. For example, for the aB phenotype,

$$f_e = \frac{3}{16} \times 2060 = 386.25 ; \quad f_o - f_e = 612 - 386.25 = + 225.75;$$

$$(f_o - f_e)^2 = (225.75)^2 = 50963.06 ; \quad \frac{(f_o - f_e)^2}{f_e} = \frac{50963.06}{386.25} = 131.94.$$

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 1515.00 \quad (\text{from Table 9.9}) ;$$

$$df = k - 1 = 4 - 1 = 3,$$

because 1 df is lost in keeping n unchanged.

The critical χ^2 values ($df = 3$) are quoted from Table C of Appendix.

$$\chi^2_{.05(3)} = 7.82 ; \quad \chi^2_{.02(3)} = 9.84 ; \quad \chi^2_{.01(3)} = 11.34 ; \quad \chi^2_{.001(3)} = 16.27.$$

As the computed χ^2 of 1515 is far higher than the critical χ^2 for 0.001 level, the computed χ^2 is highly significant ($P < 0.001$). Hence, there is a *significant difference* between the observed distribution and the proposed 9:3:3:1 distribution. So, the F_2 generation *does not obey* the Mendelian law of independent assortment.

2. χ^2 test for linkage :

This is a test of independence. The H_0 proposes here that there is no linkage between A and B (or a and b) and they are segregating independent of each other ; the H_0 contends that any departure from the frequencies, expected on that basis, arises from mere chance. To test this H_0 , χ^2 is computed as follows.

(a) The observed frequencies (f_o) are arranged in a fourfold contingency table (Table 9.10) and χ^2 is computed using the cell frequencies and marginal totals (f_r and f_c).

Table 9.10. Fourfold contingency table for linkage.

	b	B	Total (f_r)
A	725 (B)	339 (A)	1064 (A + B)
a	384 (D)	612 (C)	996 (C + D)
Total (f_c)	1109 (B + D)	951 (A + C)	2060 (n)

$$\chi^2 = \frac{n(AD - BC)^2}{(A+B)(A+C)(B+D)(C+D)} = \frac{2060 (339 \times 384 - 725 \times 612)^2}{1064 \times 951 \times 1109 \times 996} = 181.17.$$

$$df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1.$$

(b) Critical χ^2 values ($df = 1$) are quoted from Table C of Appendix.

$$\chi^2_{0.02(1)} = 5.41 ; \quad \chi^2_{0.01(1)} = 6.64 ; \quad \chi^2_{0.001(1)} = 10.83.$$

As the computed χ^2 far exceeds the critical χ^2 for the 0.001 level, the computed χ^2 is highly significant ($P < 0.001$). So, the H_0 of nonlinkage may be rejected — there is a *significant linkage* between the two relevant genes.

Example 9.2.8.

Out of a total of 288 unsuccessful candidates in an examination, 104 failed in English, 106 in Mathematics and 78 in Bengali. Test whether the results diverge significantly from the expectation that an equal proportion of unsuccessful examinees would fail in each subject ($\alpha = 0.05$).

Solution :

That an equal proportion of unsuccessful examinees should fail in each subject, is a *hypothesis of equal probability* of failure. According to it, out of a total of 288 unsuccessful examinees, $288/3$ or 96 should fail in each subject. So, for each subject, f_e amounts to 96 (Table 9.11).

Table 9.11. Computation of χ^2 for the hypothesis of equal probability.

Subject	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
English	104	96	+ 8	64	0.667
Bengali	78	96	- 18	324	3.375
Maths	106	96	+ 10	100	1.042
Total	288	288			5.084 (χ^2)

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 5.084 ; \quad df = k - 1 = 3 - 1 = 2,$$

because one of 3 classes loses its freedom to vary its f_o if n is to be kept constant. From Table C,

$$\alpha = 0.05 ; \quad \text{critical } \chi^2_{0.05(2)} = 5.99.$$

Because the computed χ^2 is lower than the critical χ^2 , the computed χ^2 is not significant ($P > 0.05$). So, the observed distribution of failed examinees in the three subjects *does not diverge significantly* from that expected on the hypothesis of equal probability.

Example 9.2.9.

In 160 samples of 5 infants each, drawn at random from a population, the number of male infants was found to have a frequency distribution (f_o) among the samples, as given in Table 9.12. Does the observed distribution differ significantly from what is expected from the assumed 1 : 1 sex-ratio of human infants ? ($\alpha = 0.05$).

Solution :

Sex being a dichotomous variable divided into two classes, binomial distribution may serve as its probability distribution.

Table 9.12. Table for computing χ^2 for the binomial distribution of male infants. ($f_e = p_e N$).

Males per sample	Number of samples (f_o)	p_e	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
5	6	0.03125	5	+ 1	1	0.20
4	23	0.15625	25	- 2	4	0.16
3	46	0.3125	50	- 4	16	0.32
2	55	0.3125	50	+ 5	25	0.50
1	27	0.15625	25	+ 2	4	0.16
0	3	0.03125	5	- 2	4	0.80
Total	160 (N)		160			2.14

(a) According to the assumed 1 : 1 sex-ratio, $p = q = 0.5$, where p and q are the proportions of males and females respectively in the population. The total number (N) of samples drawn is 160. As each sample consists of 5 infants ($n = 5$), the successive terms of the following binomial expansion of $(p + q)^n$ give the relative expected proportions or probabilities (p_e) of obtaining respectively 5, 4, 3, 2, 1 and 0 males in each sample.

$$(p + q)^n = p^n + np^{n-1}q + \frac{n(n-1)}{1 \times 2}p^{n-2}q^2 + \frac{n(n-1)(n-2)}{1 \times 2 \times 3}p^{n-3}q^3 + \dots + q^n,$$

$$\text{or, } (0.5 + 0.5)^5 = (0.5)^5 + 5 \times (0.5)^4 \times 0.5 + \frac{5 \times 4}{1 \times 2} \times (0.5)^3 \times (0.5)^2 + \frac{5 \times 4 \times 3}{1 \times 2 \times 3} \times (0.5)^2 \times (0.5)^3$$

$$+ \frac{5 \times 4 \times 3 \times 2}{1 \times 2 \times 3 \times 4} \times 0.5 \times (0.5)^4 + (0.5)^5$$

$$= 0.03125 + 0.15625 + 0.3125 + 0.3125 + 0.15625 + 0.03125.$$

Each p_e thus calculated is entered in Table 9.12.

(b) The expected frequencies (f_e) of males in the samples are then computed by multiplying the p_e of each class with the total number of samples ($N = 160$). For example, the f_e for 3 males per sample amounts to :

$$p_e N = 0.3125 \times 160 = 50 \text{ (Table 9.12).}$$

(c) The differences ($f_o - f_e$) between the observed and expected frequencies are computed and squared. These squared differences, $(f_o - f_e)^2$, thus worked out, are then used in computing χ^2 (Table 9.12).

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 2.14 ; \quad \alpha = 0.05 ; \quad df = k - 2 = 6 - 2 = 4,$$

because there are 6 classes ($k = 6$) and two df are lost in respectively keeping N unaltered and estimating the population proportion.

$$\text{Critical } \chi^2_{.05(4)} = 9.49 \quad (\text{Table C of Appendix}).$$

The computed χ^2 of 2.14 is far lower than the critical χ^2 for 0.05 level. Hence, the computed χ^2 is not significant ($P > 0.05$). So, the observed distribution *does not differ significantly* from that expected on the assumed 1 : 1 sex-ratio.

9.3 G TEST

This nonparametric test is frequently preferred to the chi square test for the *analysis of frequencies*. Like the chi square test, the G test is applicable irrespective of the normality or non-normality of distribution of the given variable in the population. It is applicable even to nominal and ordinal variables and to small samples. For large samples, the *log likelihood ratio statistic* (G) has sampling distributions closely similar to those of χ^2 . So, the computed G is compared with critical χ^2 values for testing its significance.

G test for goodness of fit

The G test is used for finding whether or not there exists a significant goodness of fit between an observed frequency distribution and a theoretically expected distribution such as the normal, binomial, Poisson, Mendelian 9:3:3:1, and equal probability distributions. The H_0 contends that the observed distribution does not differ significantly from the proposed distribution, and that any difference between the two distributions can be explained away by mere chances of random sampling. The probability P of this H_0 being correct is then worked out as follows.

(a) On the basis of the proposed hypothesis, the expected frequencies (f_e) are calculated for the different classes of the distribution in the same way as in case of the χ^2 test (§ 9.2). The computed f_e values would thus conform to the proposed distribution, viz., a normal, binomial, Poisson or Mendelian distribution, as the case may be.

(b) Using the observed frequencies (f_o) and the corresponding expected frequencies (f_e), the statistic G is computed with a calculator having the $\log_e(\ln)$ key.

$$G = 2 \sum \left(f_o \log_e \frac{f_o}{f_e} \right).$$

Yates' correction : (i) If the distribution has only 2 classes ($k = 2$) and the sample size is lower than 200 ($n < 200$), *Yates' correction* has to be applied to avoid an upward bias in the computed G . For this, 0.5 is subtracted from each f_o higher than the corresponding f_e while 0.5 is added to each f_o lower than the corresponding f_e — the Yates' correction is thus intended to bring each $(f_o - f_e)$ closer to 0 by 0.5. If Yates' correction has been applied, the corrected f_o values should be used, instead of the original f_o values, for computing G . (ii) For distributions with more than 2 classes ($k > 2$), any class with f_e less than 5 should be combined with an immediately adjacent class so as to raise their combined f_e to 5 or above.

(c) The degrees of freedom of the computed G are calculated in the same principle as in the case of χ^2 (page 210) — one df is lost for keeping n unaltered and one more df is lost for each parameter involved in the expected distribution. So, where k is the number of classes in the distribution, the df will amount to $(k - 2)$ for binomial and Poisson distributions, $(k - 3)$ for normal distributions, and $(k - 1)$ for distributions involving no parameter (e.g., Mendelian phenotype distributions, equal probability distribution, etc.)

(d) The computed G is next compared with critical χ^2 values having the computed df . If the computed G exceeds or equals the critical χ^2 for a chosen level of significance, the probability P for the H_0 being correct is considered too low and the computed G is significant ($P \leq \alpha$) — observed and expected distributions then differ significantly and there is no significant goodness of fit between the distributions. If, on the contrary, G is lower than the critical χ^2 , it is not significant ($P > \alpha$) — there is a significant goodness of fit between observed and expected distributions and they do not differ significantly.

Example 9.3.1.

Crossing of a grey-bodied scarlet-eyed *Drosophila* with a black-bodied red-eyed one produced all grey-bodied red-eyed flies in the first filial (F_1) generation. On crossing these F_1 flies, the F_2 generation gave following numbers of flies of different phenotypes : grey-bodied red-eyed : 362 ; black-bodied red-eyed : 128 ; grey-bodied scarlet-eyed : 122 ; black-bodied scarlet-eyed : 44. Apply G test to explore whether there exists a goodness of fit between the observed F_2 phenotype distribution and the Mendelian 9:3:3:1 distribution of phenotypes.

Solution :

The H_0 contends that there is no significant difference between the observed distribution of F_2 phenotypes and the proposed Mendelian 9:3:3:1 distribution. A two-tail G test is undertaken to find the probability P of this H_0 being correct.

(a) The observed frequencies (f_o) in different phenotypes are tabulated in Table 9.13.

(b) In terms of the Mendelian 9:3:3:1 distribution, the proportions of the total frequency (n) in the respective phenotypes should be :

grey-red : 9/16 ; black-red : 3/16 ; grey-scarlet : 3/16 ; black-scarlet : 1/16.

The sample size (n) is multiplied by these proportions to give the respective f_e values of the phenotypes. For example, for the black-red phenotype,

$$n = 656 \text{ (from Table 9.13) ; } f_e = n \times \frac{3}{16} = 656 \times \frac{3}{16} = 123.$$

(c) Using a calculator, the f_o/f_e , $\ln(f_o/f_e)$ and $f_o \times \ln(f_o/f_e)$ values are worked out for each phenotype and entered in Table 9.13. For example, for the black-red phenotype,

$$f_o/f_e = \frac{128}{123} = 1.0407 ; \quad \ln(f_o/f_e) = \ln(1.0407) = 0.0399 ;$$

$$f_o \times \ln(f_o/f_e) = 128 \times 0.0399 = 5.1072.$$

Table 9.13. Table for computing G for goodness of fit with Mendelian distribution.

Phenotypes	f_o	f_e	f_o/f_e	$\ln(f_o/f_e)$	$f_o \times \ln(f_o/f_e)$
Grey-red	362	369	0.9810	- 0.0192	- 6.9504
Black-red	128	123	1.0407	0.0399	5.1072
Grey-scarlet	122	123	0.9919	- 0.0081	- 0.9882
Black-scarlet	44	41	1.0732	0.0706	3.1064
Total	656 (n)	656			0.2750

$$\therefore G = 2 \sum [f_o \ln(f_o/f_e)] = 2 \times 0.2750 = 0.55.$$

Because the proposed distribution is a Mendelian distribution, $df = k - 1 = 4 - 1 = 3$.

Critical χ^2 values ($df = 3$) are quoted from Table C of Appendix,

$$\chi^2_{0.01(3)} = 11.34 ; \quad \chi^2_{0.05(3)} = 7.82 ; \quad \chi^2_{0.10(3)} = 6.25.$$

As the computed G is lower than the critical χ^2 for 0.05 level, the probability P of the H_0 being correct is considered too high and H_0 cannot be rejected ($P > 0.05$). So, the computed G is considered not significant and it is inferred that there is no significant difference between the observed phenotype distribution and the Mendelian 9:3:3:1 distribution ; in other words, there is a significant goodness of fit between the two distributions.

Example 9.3.2.

Find whether or not the observed distribution of serum iron concentrations, presented in *Example 6.4.1*, differs significantly from the normal distribution.

Solution :

(a) Expected frequencies (f_e) of a best-fitting normal distribution are first computed. In the present case, that has already been done stepwise in *Example 6.4.1* and presented in Table 6.4 (page 84). The observed and expected frequencies of that table are entered in Table 9.14 for further computations.

Table 9.14. Table for computing G for goodness of fit with best-fitting normal distribution.

Classes	f_o	f_e	f_o/f_e	$\ln(f_o/f_e)$	$f_o \times \ln(f_o/f_e)$
100-109	6	3.5	1.428571	0.356675	6.063475
110-119	11	8.4			
120-129	10	14.7	0.680272	- 0.385263	- 3.852630
130-139	17	18.4	0.923913	- 0.079137	- 1.345329
140-149	16	16.4	0.975610	- 0.024692	- 0.395072
150-159	13	10.5	1.307190	0.267880	5.357600
160-169	7	4.8			
Total	80 (n)				5.828044

(b) Two of the seven classes, viz., 100-109 and 160-169, have f_e scores less than 5. Each of these two classes is, therefore, combined with the next class. The final number of classes thus comes to five only : $k = 5$.

(c) Using a calculator, f_o/f_e , $\ln(f_o/f_e)$ and $f_o \ln(f_o/f_e)$ values are calculated for each class and entered in Table 9.14. The sum of all the $f_o \ln(f_o/f_e)$ scores is then used in computing G .

$$G = 2 \sum \left(f_o \ln \frac{f_o}{f_e} \right) = 2 \times 5.828044 = 11.66.$$

Because the proposed distribution is a normal distribution, $df = k - 3 = 5 - 3 = 2$.

Critical χ^2 scores :

$$\chi^2_{0.05(2)} = 5.99 ; \quad \chi^2_{0.02(2)} = 7.82 ; \quad \chi^2_{0.01(2)} = 9.21 ; \quad \chi^2_{0.001(2)} = 13.82.$$

As the computed G exceeds the critical χ^2 for 0.01 level, $P < 0.01$. Thus the H_0 is rejected and the computed G is significant. So, there is a *significant difference* between the observed distribution and the normal distribution.

G test of independence or association

G test can be used to find whether or not two variables have significant association with each other. The H_0 contends here that there is no significant association between the variables and they are independent of each other.

(a) A contingency table is constructed arranging the classes of one variable along its rows, and those of the other variable along the columns. This gives the $r \times c$ -fold contingency table with r number of rows for one variable and c number of columns for the other

variable. The observed frequencies (f_o) of cases, belonging to specific combinations of classes of two variables, are then entered in the respective cells of the contingency table (Table 9.15). The total of f_o scores of cells of each row or column is entered at the margin of the table as the marginal total, f_r or f_c respectively, of that row or column.

(b) The f_o , f_r and f_c values are used in computing G with a calculator having the \log_e (ln) key,

$$G = 2 \left[\sum (f_o \ln f_o) - \left(\sum (f_r \ln f_r) + \sum (f_c \ln f_c) \right) + n \ln n \right].$$

$$df = (r - 1)(c - 1).$$

Yates' correction : In case of the 2×2 -fold contingency tables, Yates' correction should be applied if the sample size is lower than 200 ($n < 200$) (Example 9.3.4). For this, 0.5 is added to the f_o values of cells B and C of the

contingency table, and is subtracted from the f_o values of cells A and D each, if the product of f_o values of A and D exceeds that of B and C, i.e., if $AD - BC$ is positive (Table 9.17). On the contrary, if $AD - BC$ is negative, 0.5 is subtracted from the f_o values of B and C, and added to those of A and D. The corrected f_o values are then used, instead of the originally observed f_o values, in computing G by the preceding formula.

(c) The computed G is next compared with critical χ^2 values having the given df . If the computed G exceeds or equals the critical χ^2 for the chosen level of significance, the P of the H_0 being correct is considered too low ($P \leq \alpha$) — the variables under consideration then possess a significant association with each other. If, on the contrary, the computed G is lower than the critical χ^2 , the variables do not have any significant association with each other and are independent of each other ($P > \alpha$).

Example 9.3.3.

40 out of 100 diabetics and 25 out of 120 nondiabetics suffered from high blood pressure (hypertension) while the others of both groups were free from hypertension. Is there a significant association between diabetes and hypertension ?

Solution :

A two-tail G test is undertaken to find the probability P of correctness of the H_0 proposing independence of the variables.

(a) The data are arranged in a 2×2 -fold contingency table (Table 9.15). The cell frequencies (f_o) of each row are totalled to give the f_r of that row ; the f_o values of each column are similarly totalled to give the f_c of that column. $n = \sum f_r = \sum f_c = 220$.

Table 9.15. Fourfold contingency table for hypertension data.

	Nonhypertensive (f_o)	Hypertensive (f_o)	Total (f_r)
Diabetic	60 (B)	40 (A)	100 (A + B)
Nondiabetic	95 (D)	25 (C)	120 (C + D)
Total (f_c)	155 (B + D)	65 (A + C)	220 (n)

(b) Each of f_o , f_r , f_c and n values is converted to \log_e value using a calculator with an ln key, and entered in Table 9.16.

(c) These values are then used in computing G .

$$\begin{aligned}
 G &= 2 \left[\Sigma(f_o \ln f_o) - \{\Sigma(f_r \ln f_r) + \Sigma(f_c \ln f_c)\} + n \ln n \right] \\
 &= 2 \left[60 \times 4.09434 + 40 \times 3.68888 + 95 \times 4.55388 + 25 \times 3.21888 \right. \\
 &\quad \left. - \{(100 \times 4.60517 + 120 \times 4.78749) + (155 \times 5.04343 + 65 \times 4.17439)\} + 220 \times 5.39363 \right] \\
 &= 9.644.
 \end{aligned}$$

$$df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1.$$

Critical χ^2 values ($df = 1$) are quoted from Table C of *Appendix*.

$$\chi^2_{0.01(1)} = 10.83; \quad \chi^2_{0.05(1)} = 3.84; \quad \chi^2_{0.1(1)} = 2.71.$$

As the computed G is higher than the critical χ^2 for 0.01 level, the probability P of the H_0 being correct is considered too low. So, there is a significant association between diabetes and hypertension ($P < 0.01$).

Table 9.16. Table for G test of independence of diabetes and hypertension.

	Nonhypertensive		Hypertensive		f_r	$\ln f_r$
	f_o	$\ln f_o$	f_o	$\ln f_o$		
Diabetic	60	4.09434	40	3.68888	100	4.60517
Nondiabetic	95	4.55388	25	3.21888	120	4.78749
f_c	155		65		220 (n)	
$\ln f_c$	5.04343		4.17439		5.39363 ($\ln n$)	

Example 9.3.4.

20 out of 48 normal persons and 37 out of 57 neurotics gave positive responses to a test item of neurotic questionnaire while others of each group gave negative responses to that test item. Is there significant association between neurotic state and response to the item?

Solution :

A two-tail G test is undertaken to find the probability P of correctness of the H_0 which contends that the test item has no association with neurotic state.

(a) The data are arranged in the cells, A, B, C and D, of a 2×2 -fold contingency table (Table 9.17). The cell frequencies (f_o) of each row are totalled to give the f_r of that row. The f_o values of each column are similarly totalled to give the f_c of that column.

Table 9.17. 2×2 -fold contingency table and corrected f_o scores for neurotic questionnaire data.

	Negative response		Positive response		Total (f_r)
	f_o	Corrected f_o	f_o	Corrected f_o	
Neurotic	20 (B)	20.5	37 (A)	36.5	57
Normal	28 (D)	27.5	20 (C)	20.5	48
Total (f_c)	48		57		105 (n)

(b) Because it is a 2×2 -fold contingency table with $n < 200$, Yates' correction is applied to all the f_o values. Here, the product of f_o scores of cells A and D is seen to exceed that of cells B and C, i.e., $AD > BC$. So, 0.5 is subtracted from the f_o of each of the cells A and D while 0.5 is added to the f_o of each of the cells B and C to get the corrected f_o scores (Table 9.17). These corrected f_o scores are used for the subsequent computations.

(c) The \log_e values of n and of each of the f_r , f_c and corrected f_o are worked out with a calculator and entered in Table 9.18. These values are then used in computing G .

Table 9.18. Table for G test of association between response to test item and neurotic state, using corrected f_o values.

	Negative response		Positive response		f_r	$\ln f_r$
	f_o	$\ln f_o$	f_o	$\ln f_o$		
Neurotic	20.5	3.02042	36.5	3.59731	57	4.04305
Normal	27.5	3.31419	20.5	3.02042	48	3.87120
f_c	48		57		105 (n)	
$\ln f_c$	3.87120		4.04305		4.65396 ($\ln n$)	

$$\begin{aligned}
 G &= 2 \left[\Sigma(f_o \ln f_o) - \{\Sigma(f_r \ln f_r) + \Sigma(f_c \ln f_c)\} + n \ln n \right] \\
 &= 2 \left[20.5 \times 3.02042 + 36.5 \times 3.59731 + 27.5 \times 3.31419 + 20.5 \times 3.02042 \right. \\
 &\quad \left. - \{(57 \times 4.04305 + 48 \times 3.87120) + (48 \times 3.87120 + 57 \times 4.04305)\} + 105 \times 4.65396 \right] \\
 &= 4.804.
 \end{aligned}$$

$$df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1.$$

Critical χ^2 values ($df = 1$) are quoted from Table C of Appendix.

$$\chi^2_{0.01(1)} = 6.64 ; \quad \chi^2_{0.02(1)} = 5.41 ; \quad \chi^2_{0.05(1)} = 3.84.$$

As the computed G exceeds the critical χ^2 for 0.05 level, it is considered significant below that level. So, there is a significant association between neurotic state and response to the given test item ($P < 0.05$).

9.4 WILCOXON SIGNED RANK TEST

This is a nonparametric alternative to Student's t test for finding the significance of difference between paired observations of single-group and matched-pair experiments. It can be applied to both continuous and discrete variables irrespective of the normality or non-normality of their distribution in the population, and even to small samples. For its application, however, it should be justifiable to assume that each pair of scores or observations occurs at random in the sample, independent of all other pairs. This test is less powerful than the t test.

The H_0 contends that there is no significant

difference between the scores of each pair of observations, and that any observed difference results from mere chances of random sampling. The probability P of this H_0 being correct is found out in the following way.

Computation :

(a) The difference between the scores of each pair of observations is worked out. All the non-zero differences are then assigned ranks in an ascending order, according to their absolute magnitudes but ignoring their algebraic signs. If two or more differences have the same absolute value, irrespective of their algebraic signs, each is assigned an average rank which is the mean

of the actual ranks that would have been assigned to them if they were not identical but consecutive. In Table 9.19, for example, the differences +2.4 and -2.4 are both considered simply as 2.4 and each of them is assigned the average rank of 9.5 which is the mean of the ranks 9 and 10 that they would have occupied if they differed in their absolute values. In such cases, the difference having the next higher absolute magnitude is given that rank which it would have occupied if the difference immediately preceding it held a separate individual rank instead of an average rank.

(b) The original algebraic signs of the differences are next assigned to their respective ranks.

(c) The ranks with identical algebraic signs are then added together. This gives two *sums of ranks* for positive and negative ranks respectively. The smaller of these rank sums, in absolute value irrespective of signs, is taken as the statistic T .

Significance :

The significance of the computed T is found out in the following ways. The closer are the values of the absolute rank sums, i.e., the larger the absolute value T of the smaller rank sum, the higher is the probability P of the H_0 being correct.

(i) For *small samples* with 50 or lower *non-zero differences* between the paired scores ($n \leq 50$), the computed T is compared with the critical T values (T_0) for the given n and for different levels of significance (Table F of Appendix). The significance level of the computed T is given by that of the critical T_0

value which equals or exceeds the computed T — the lower the latter, the higher is its significance. In other words, the observed difference between the sample means is considered significant ($P \leq \alpha$) only when the computed T is *equal to or lower than* the T_0 for the chosen level of significance or probability ($T_0 \geq T$). There is no significant difference between the sample means ($P > \alpha$) if T exceeds T_0 . The significance level for a *one-tail test* is half that for a two-tail test.

(ii) For *large samples* with more than 25 *non-zero differences* ($n > 25$), the computed T is transformed into Student's t score, using the mean sum of ranks (\bar{T}) and the SE ($s_{\bar{T}}$).

$$\bar{T} = \frac{n(n+1)}{4}; \quad s_{\bar{T}} = \sqrt{\frac{n(n+0.5)(n+1)}{12}};$$

$$t = \frac{T - \bar{T}}{s_{\bar{T}}}; \quad df = \infty.$$

The computed t is next compared with critical t scores ($df = \infty$) for different levels of significance (Table B of Appendix). The observed difference between the sample means is considered significant only when the computed t *exceeds or equals* the critical t for the chosen level of significance ($P \leq \alpha$).

Inaccuracies :

The signed rank test suffers from two inaccuracies : (i) average ranks are used for tied differences instead of their true ranks ; (ii) the differences given consecutive ranks may differ from each other to different extents in magnitude. In Table 9.19, for example, differences ranked as 6 and 7 differ in magnitude by 0.6 while those ranked as 7 and 8 differ by 0.1.

Example 9.4.1.

Blood hemoglobin values (g per dL) of ten pernicious anemia patients, respectively before and after vitamin B_{12} therapy, are given in the first three columns of Table 9.19. Find whether or not the therapy has produced significant changes in the hemoglobin values.

Solution :

A two-tail signed rank test is undertaken.

(a) The data are tabulated in Table 9.19. The difference ($X_1 - X_2$) between the scores of each pair of observations is worked out and entered in the table. The non-zero differences are assigned ranks in an ascending order according to their absolute values, ignoring their signs. Average ranks are given to the tied differences ; e.g., the tied differences, +2.4 and -2.4, are each assigned the average rank of 9.5, ignoring their signs.

(b) The original algebraic signs of the differences are then assigned to their respective ranks ; e.g., the differences, + 2.4 and - 2.4, each given the average rank of 9.5, now bear the signed ranks of + 9.5 and - 9.5, respectively.

Table 9.19. Signed rank test of hemoglobin data.

Patient	Hemoglobin (g dL ⁻¹)		Differences ($X_1 - X_2$)	Signed ranks
	before therapy (X_1)	after therapy (X_2)		
1	12.0	14.2	-2.2	-8
2	12.4	14.5	-2.1	-7
3	12.2	13.7	-1.5	-6
4	14.6	12.2	+2.4	+9.5
5	11.5	13.9	-2.4	-9.5
6	10.8	11.3	-0.5	-3.5
7	12.6	13.1	-0.5	-3.5
8	13.1	14.2	-1.1	-5
9	10.9	10.8	+0.1	+1
10	10.3	10.0	+0.3	+2

(c) The ranks with identical algebraic signs are added to give two sums of ranks, omitting the algebraic signs of the sums.

Absolute sum of positive ranks : $9.5 + 1 + 2 = 12.5$.

Absolute sum of negative ranks : $8 + 7 + 6 + 9.5 + 3.5 + 3.5 + 5 = 42.5$.

$T =$ the smaller sum of ranks = 12.5.

(d) The computed T is compared with two-tail critical T_0 values for the n number of non-zero differences in the sample ($n = 10$). From Table F of Appendix,

α 0.02 : $T_0 = 5$; α 0.05 : $T_0 = 8$; α 0.10 : $T_0 = 11$.

As the computed T of 12.5 exceeds the critical T_0 for even 0.05 level of significance, the computed T is not significant. Hence, there is no significant difference between the hemoglobin values before and after therapy with B_{12} ($P > 0.05$).

Example 9.4.2.

Mean corpuscular hemoglobin values (MCH in pg) of 26 macrocytic anemia patients, respectively before and after folate therapy, are presented in the 2nd and 3rd columns of Table 9.20. Find whether or not the therapy has produced significant changes in the MCH values.

• Solution :

The H_0 contends that the MCH values do not differ significantly between the two groups. A two-tail signed rank test is undertaken to find the probability P of correctness of this H_0 .

(a) The difference ($X_1 - X_2$) between the scores of each pair is worked out and entered in the table. The non-zero differences are assigned ranks in an ascending order according to their absolute values, ignoring their signs. Average ranks are given to the tied differences ; e.g., the tied differences, viz, +2, +2 and -2, are each given the average rank of 3 ignoring their signs.

(b) The original algebraic signs of the differences are then assigned to their respective ranks ; e.g. the differences, viz., +2, +2 and -2, each given the average rank of 3, now bear the signed ranks of +3, +3 and -3, respectively.

Table 9.20. Signed rank test of MCH data.

Patient	MCH values		$X_1 - X_2$	Signed ranks
	before (X_1)	after (X_2)		
1	34	31	+ 3	+ 5
2	38	30	+ 8	+ 16.5
3	27	33	- 6	- 10.5
4	40	30	+ 10	+ 22
5	35	25	+ 10	+ 22
6	32	30	+ 2	+ 3
7	30	28	+ 2	+ 3
8	41	34	+ 7	+ 13.5
9	28	30	- 2	- 3
10	24	10	+ 14	+ 25
11	35	31	+ 4	+ 6
12	36	30	+ 6	+ 10.5
13	35	25	+ 10	+ 22
14	30	29	+ 1	+ 1
15	36	26	+ 10	+ 22
16	25	30	- 5	- 7.5
17	22	30	- 8	- 16.5
18	41	34	+ 7	+ 13.5
19	32	24	+ 8	+ 16.5
20	38	30	+ 8	+ 16.5
21	38	32	+ 6	+ 10.5
22	40	34	+ 6	+ 10.5
23	28	10	+ 18	+ 26
24	15	25	- 10	- 22
25	31	22	+ 9	+ 19
26	28	33	- 5	- 7.5

(c) The ranks bearing identical signs are added to give two rank sums, omitting the algebraic signs of these sums.

Absolute sum of negative ranks : $10.5 + 3 + 7.5 + 16.5 + 22 + 7.5 = 67.$

Absolute sum of positive ranks : $5 + 16.5 + 22 + 22 + 3 + 3 + 13.5 + 25 + 6 + 10.5 + 22 + 1 + 22 + 13.5 + 16.5 + 16.5 + 10.5 + 10.5 + 26 + 19 = 284.$

$T =$ the smaller rank sum $= 67.$

$$\bar{T} = \frac{n(n+1)}{4} = \frac{26(26+1)}{4} = 175.5.$$

$$s_{\bar{T}} = \sqrt{\frac{n(n+0.5)(n+1)}{12}} = \sqrt{\frac{26(26+0.5)(26+1)}{12}} = 39.37.$$

$$t = \frac{T - \bar{T}}{s_{\bar{T}}} = \frac{67 - 175.5}{39.37} = -2.756, \text{ or } 2.756. \quad df = \infty.$$

Critical t scores (from Table B of Appendix) :

$$t_{.05(\infty)} = 1.960 ; t_{.02(\infty)} = 2.326 ; t_{.01(\infty)} = 2.576 ; t_{.001(\infty)} = 3.291.$$

As the computed t of 2.756 (ignoring its sign) is higher than the critical t of 0.01 level, the computed t is significant below the 0.01 level. So, there is a significant difference between the MCH values of the two groups — the therapy has produced *significant changes* in MCH ($P < 0.01$).

9.5 WILCOXON COMPOSITE RANK TEST OR RANK SUM TEST

This is a fairly good nonparametric alternative to Student's t test for finding the *significance of difference between unpaired observations of two independent groups*. It can be applied to both continuous and discrete variables, irrespective of their normal or non-normal distributions in the population, and even to small samples. But for its application, it should be justifiable to assume that each score occurs at random and independent of all other scores. It is somewhat less powerful than the t test, but has a higher efficiency than the t test in case of non-normal distributions.

The H_0 proposes that there is no significant difference between the scores of the two independent groups, and that any observed difference has resulted from mere chances of random sampling. The probability P of this H_0 being correct is found out as follows.

Computation :

(a) All the scores of both the groups, taken

together as a *composite group*, are ranked in a single ascending order of their values. Two or more identical or tied scores, whether belonging to the same group or to separate groups, are each assigned an average rank which is the arithmetic mean of the actual ranks that would have been given to those scores if they were consecutive scores instead of being identical. The score next higher than those of a tied set is assigned the same rank as it would have got if the tied scores had separate consecutive ranks instead of an average rank.

(b) The ranks of the scores of each group are next totalled separately. This gives two sums of ranks, T_1 and T_2 , for the two groups. As the H_0 contends that there is no significant difference between the scores of the two groups, the ranks should be evenly distributed between the groups to make T_1 equal to T_2 if the H_0 is correct.

Significance :

The significance of the rank sums is explored in one of the following ways.

(i) For *small groups of unequal sizes* ($n_1 = 3$ to 25, $n_2 = 3$ to 50), the sum of ranks of the *smaller group* is taken as the statistic T and compared with the critical upper and lower T values (T_u and T_l respectively) for the given combination of group sizes (N, M) and the chosen level of significance (Table G of Appendix). If the computed T is found to lie between T_u and T_l , i.e., $T_u > T > T_l$, the H_0 cannot be rejected and the means of the two groups do not differ significantly ($P > \alpha$). If T_l exceeds or equals the computed T , i.e., $T_l \geq T$, the mean of the smaller group is significantly smaller than that of the larger one. If T exceeds or equals T_u , i.e., ($T \geq T_u$), the mean of the smaller group is significantly higher than that of the larger one.

(ii) For *small groups of equal size* ($n_1 = n_2 \leq 25$), the smaller of the two sums of ranks is taken as the statistic T and compared with the critical lower T value (T_l) for the given combination of group sizes and the level of significance chosen for the two-tail or one-tail test as the case may be. If $T_l \geq T$, there is a significant difference between the scores of the two groups ($P \leq \alpha$).

(iii) For *large groups of equal size* ($n_1 = n_2 > 20$), the critical T_α values are computed for

different levels of significance, using the mean rank sum (\bar{T}), its SE ($s_{\bar{T}}$), and the total size N of the two groups ($N = n_1 + n_2$).

$$\bar{T} = \frac{T_1 + T_2}{2} = \frac{N(N+1)}{4}; \quad s_{\bar{T}} = \sqrt{\frac{NT}{12}}$$

For *two-tail tests*, T_α values are given by: $T_\alpha = \bar{T} - t_{\alpha/2} s_{\bar{T}}$; thus, for 0.05, 0.01 and 0.001 significance levels,

$$\begin{aligned} T_{.05} &= \bar{T} - 1.96 s_{\bar{T}}; & T_{.02} &= \bar{T} - 2.33 s_{\bar{T}}; \\ T_{.01} &= \bar{T} - 2.58 s_{\bar{T}}; & T_{.001} &= \bar{T} - 3.29 s_{\bar{T}}. \end{aligned}$$

Similarly, for *one-tail tests*,

$$\begin{aligned} T_{.05} &= \bar{T} - 1.65 s_{\bar{T}}; & T_{.025} &= \bar{T} - 1.96 s_{\bar{T}}; \\ T_{.01} &= \bar{T} - 2.33 s_{\bar{T}}; & T_{.005} &= \bar{T} - 2.58 s_{\bar{T}}. \end{aligned}$$

The H_0 is rejected and the observed difference is considered significant if the computed T , i.e., the smaller sum of ranks, is *equal to or lower than* the T_α for the chosen significance level.

Inaccuracies :

This test suffers from two inaccuracies : (i) average ranks are used for tied scores instead of their true ranks ; (ii) no consideration is given to the fact that the scores having consecutive ranks may differ from each other to different extents in magnitude.

Example 9.5.1.

The mean corpuscular hemoglobin (MCH) values in picograms, estimated in 9 normal men and 11 macrocytic anemia patients, are given in the first and third columns respectively of Table 9.21. Is the MCH significantly lower in normal persons than in macrocytic anemia patients?

Solution :

The H_0 contends that the MCH is not significantly lower in normal men than in macrocytic anemia patients. A *one-tail composite rank test* is undertaken to find the probability P of this H_0 being correct.

(a) Ranks are assigned in an ascending order to the MCH values of both groups taken together ; average ranks are given to tied values (Table 9.21). For example, each of two tied scores of 28, one occurring in each group, is given an average rank of 4.5 which is the mean of the separate ranks 4 and 5 that they would have got if they were not tied.

Table 9.21. Table for composite rank test of MCH data.

Normal		Macrocytic anemia	
Scores	Ranks	Scores	Ranks
31	9	35	13
30	7	40	17
25	2	37	14
30	7	33	11.5
26	3	38	15.5
28	4.5	30	7
24	1	41	18
32	10	38	15.5
33	11.5	28	4.5
		42	19
		45	20
Total	55 (T_1)		155 (T_2)

(b) The ranks of each group are totalled separately to give the respective rank sums : $T_1 = 55$; $T_2 = 155$. As the two groups differ in size, the rank sum T_1 of the smaller group ($n_1 = 9$) is taken as the statistic T . $\therefore T = 55$.

(c) The computed T is compared with one-tail critical T_u and T_l values for the given combination of group sizes, viz., $n_1 = 9$ and $n_2 = 11$ (Table G of Appendix).

$$\alpha 0.005 : T_u = 128 ; T_l = 61.$$

$$\alpha 0.025 : T_u = 121 ; T_l = 68.$$

Because even at 0.005 level, the computed T of 55 is lower than the T_l value of 61 ($T_l > T$), the P of the H_0 being correct is considered too low. So, the mean of the smaller group (normal) is significantly lower than that of the larger group (anemic) ($P < 0.005$).

Example 9.5.2.

The strengths of the kneejerk reflex (degrees of arc), measured under relaxed condition in 10 athletes and 10 nonathletes respectively, are tabulated in the first and third columns of Table 9.22. Does the mean strength of the kneejerk reflex differ significantly between the groups ?

Solution :

A two-tail composite rank test is applied.

(a) Ranks are assigned in an ascending order to the scores of both groups taken together ; average ranks are assigned to tied scores (Table 9.22). For example, each of two tied scores of 35, one occurring in each group, is given the average rank of 18.5 which is the mean of the ranks 18 and 19 that they would have got if they were not tied.

(b) The ranks of each group are totalled separately to give the respective rank sums : $T_1 = 129$; $T_2 = 81$. As the samples are of equal size ($n_1 = n_2 = 10$), the smaller sum of ranks, viz., T_2 , is taken as the statistic T . $\therefore T = 81$.

(c) Two-tail critical T_1 values ($n_1 = n_2 = 10$) for different significance levels are quoted from Table G of Appendix.

$$\alpha \ 0.01 : T_1 = 71 ; \quad \alpha \ 0.05 : T_1 = 79.$$

Table 9.22. Table for composite rank test of kneejerk data.

Athletes		Nonathletes	
Kneejerk strengths	Ranks	Kneejerk strengths	Ranks
31	16.5	35	18.5
30	14.5	26	9.5
22	7	14	2.5
30	14.5	20	5
26	9.5	14	2.5
28	13	21	6
19	4	31	16.5
35	18.5	27	11.5
37	20	24	8
27	11.5	13	1
Total	129 (T_1)		81 (T_2)

Because $T > T_1$ even at 0.05 level, the P of the H_0 being correct is considered too high. So, there is no significant difference between the mean kneejerk strengths of the two groups ($P > 0.05$).

Example 9.5.3.

The memory test scores of two groups of students are given in the first and third columns, respectively, of Table 9.23. Is there any significant difference between the mean memory scores of the two groups ?

Solution :

A two-tail composite rank test is undertaken.

(a) Ranks are assigned in an ascending order to the scores of both the groups taken together, giving average ranks to the tied scores (Table 9.23).

(b) The ranks of each group are totalled separately to give the respective sums of ranks : $T_1 = 587.5$; $T_2 = 315.5$. The smaller of the two rank sums, viz., T_2 , is taken as the statistic T because the groups are of equal size ($n_1 = n_2 = 21$). $\therefore T = 315.5$.

Table 9.23. Table for composite rank test of memory scores.

Group A		Group B	
Memory scores (X_1)	Ranks	Memory scores (X_2)	Ranks
20	15.5	21	17.5
33	33.5	18	12
35	37	17	10
34	35	7	1
25	21.5	9	3
38	41	18	12
25	21.5	20	15.5
27	23.5	23	19
24	20	35	37
31	29.5	31	29.5
33	33.5	16	9
37	40	14	8
32	31.5	11	4
30	28	12	5.5
29	26	12	5.5
19	14	27	23.5
18	12	35	37
21	17.5	32	31.5
29	26	13	7
36	39	29	26
39	42	8	2
Total	587.5 (T_1)		315.5 (T_2)

(c) Because the samples are large ($n_1 = n_2 > 20$), the two-tail critical T_α values are computed as follow for different levels.

$$N = n_1 + n_2 = 21 + 21 = 42. \quad \bar{T} = \frac{N(N+1)}{4} = \frac{42 \times 43}{4} = 451.5.$$

$$s_{\bar{T}} = \sqrt{\frac{N\bar{T}}{12}} = \sqrt{\frac{42 \times 451.5}{12}} = 39.75.$$

$$T_{.05} = \bar{T} - 1.96s_{\bar{T}} = 451.5 - 1.96 \times 39.75 = 373.6 ;$$

$$T_{.01} = \bar{T} - 2.58s_{\bar{T}} = 451.5 - 2.58 \times 39.75 = 348.9 ;$$

$$T_{.001} = \bar{T} - 3.29s_{\bar{T}} = 451.5 - 3.29 \times 39.75 = 320.7.$$

Because the computed T of 315.5 is lower than even $T_{.001}$, the computed T is significant below the 0.001 level. So, there is a significant difference between the mean scores of the groups ($P < 0.001$).

Example 9.5.4.

The grip strengths (kg) of 21 athletes and 21 nonathletes are given in respectively the 1st and 3rd columns of Table 9.24. Is there is any significant difference between the grip strengths of the two groups ?

Solution :

A two-tail composite rank test is undertaken.

(a) Ranks are assigned in an ascending order to the grip strengths of both the groups taken together, giving average ranks to the tied values (Table 9.24).

Table 9.24. Composite rank test of grip strength data.

Athletes		Nonathletes	
Grip strengths	Ranks	Grip strengths	Ranks
10	18	6	5
12	25	4	2
9	14	3	1
16	37	6	5
11	22	7	8
17	39	8	11
12	25	10	18
14	31	5	3
15	34	11	22
13	28	9	14
18	40.5	13	28
8	11	6	5
11	22	9	14
19	42	8	11
16	37	12	25
10	18	14	31
7	8	10	18
14	31	10	18
15	34	15	34
18	40.5	7	8
16	37	13	28
Total	594 (T_1)		309 (T_2)

(b) The ranks of each group are *totalled separately* to give the respective rank sums (Table 9.24). Because the samples are of equal size, the smaller rank sum is taken as the statistic T .

$$T_1 = 594 ; \quad T_2 = 309 ; \quad n_1 = n_2 = 21 ; \quad N = n_1 + n_2 = 21 + 21 = 42 ;$$

$$T = \text{smaller rank sum} = 309.$$

(c) Because the groups are large (> 20), two-tail critical T_{α} values are computed as follows for different levels.

$$\bar{T} = \frac{T_1 + T_2}{2} = \frac{594 + 309}{2} = 451.5 ; \quad s_{\bar{T}} = \sqrt{\frac{NT}{12}} = \sqrt{\frac{42 \times 451.5}{12}} = 39.75.$$

$$T_{.05} = \bar{T} - 1.96s_{\bar{T}} = 451.5 - 1.96 \times 39.75 = 373.59 ;$$

$$T_{.02} = \bar{T} - 2.33s_{\bar{T}} = 451.5 - 2.33 \times 39.75 = 358.88 ;$$

$$T_{.01} = \bar{T} - 2.58s_{\bar{T}} = 451.5 - 2.58 \times 39.75 = 348.95 ;$$

$$T_{.001} = \bar{T} - 3.29s_{\bar{T}} = 451.5 - 3.29 \times 39.75 = 320.72.$$

As the computed T of 309 is lower than even $T_{.001}$, the computed T is significant below the 0.001 level. So, there is a significant difference between the mean grip strengths of athletes and nonathletes ($P < 0.001$).

9.6 MANN-WHITNEY U TEST

This is an efficient nonparametric alternative to Student's t test. Its power is only slightly lower than that of the t test, but is higher than that of the composite rank test or the median test. It can be applied to both continuous and discrete measurement variables, irrespective of normality or non-normality of their distributions in the population, and also to very small samples. It should, however, be justifiable to assume that each score occurs in the sample at random and independent of all other scores. It is particularly used to test the significance of differences between *unpaired observations of two independent groups of unequal sizes* ($n_1 \neq n_2$). It is worked out as follows.

(a) Ranks are assigned in an ascending order to all the scores of both the groups taken together, giving average ranks to tied scores. Like all other rank tests, the U test suffers from two inaccuracies due to (i) the assignment of average ranks to tied scores, and (ii) the ranking of scores ignoring the magnitudes of differences between them.

(b) The ranks of each group are totalled separately to give the respective sums of ranks, R_1 and R_2 .

(c) Either of the rank sums may be used in computing the Mann-Whitney statistic U .

$$U_1 = n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1 ;$$

$$U_2 = n_1n_2 + \frac{n_2(n_2+1)}{2} - R_2.$$

(d) The H_0 contends that there is no significant difference between the scores of the two groups and they belong to the same population. This leads to the proposition that R_1 and R_2 are identical, any observed difference between them being due to mere chances of random sampling. The probability of this H_0 being correct may be worked out in two alternative ways.

(i) For small groups : If any of the groups consists of less than 8 cases, the smaller of the two statistics, U_1 or U_2 , is compared with the critical U values at chosen α levels and for the specific combination of n_1 and n_2 (Table E of Appendix). Critical one-tail and two-tail U values are used for respectively one-tail and two-tail Mann-Whitney tests. Only if the smaller computed U is equal to or lower than the critical U at the chosen α , the H_0 is rejected and the two means are considered to differ significantly.

(ii) *For large groups* : If both the groups consist of 8 or more cases, the value U_e , as expected from the H_0 , is first computed.

$$U_e = \frac{n_1 n_2}{2}.$$

The z score is then computed from any of the two U values worked out from the respective sums of ranks. Where s_U is the SE of U ,

$$s_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}};$$

$$z = \frac{U_1 - U_e}{s_U}; \quad \text{or, } z = \frac{U_2 - U_e}{s_U}.$$

The two z scores would have the same

absolute value, but one of them would bear the positive sign, and the other the negative sign. Ignoring the algebraic sign of the computed z scores, the probability P of the H_0 being correct is calculated using the unit normal curve areas (Table A of Appendix). For a two-tail test,

$$P = 2 [0.5000 - (\text{fractional area of unit normal curve from } \mu \text{ to the computed } z)].$$

For a one-tail test,

$$P = 0.5000 - (\text{fractional area of unit normal curve from } \mu \text{ to the computed } z).$$

The difference between the group means is considered significant only if P is equal to or lower than the chosen α ($P \leq \alpha$).

Example 9.6.1.

The winglengths (mm) of two samples of houseflies are given below.

Sample 1 : 3.9, 4.3, 4.7, 3.7, 4.2, 4.1, 4.8, 5.3, 4.9, 5.2, 5.5.

Sample 2 : 4.8, 3.6, 4.5, 3.9, 4.6, 4.0, 3.8.

Is there a significant difference between the means of the two samples ? ($\alpha = 0.02$).

Solution :

A two-tail Mann-Whitney U test is done.

(a) The scores of both samples taken together are assigned ranks in an ascending order, giving average ranks to the tied scores (Table 9.25).

(b) The ranks of each sample are totalled separately to give the respective sums of ranks, R_1 and R_2 .

$$R_1 = 122; \quad R_2 = 49.$$

(c) The statistic U is computed from both rank sums, R_1 and R_2 .

$$n_1 = 11; \quad n_2 = 7.$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 = 11 \times 7 + \frac{11 \times 12}{2} - 122 = 21.$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 = 11 \times 7 + \frac{7 \times 8}{2} - 49 = 56.$$

Table 9.25. U test of winglength data.

Sample 1		Sample 2	
Winglengths (X_1)	Ranks	Winglengths (X_2)	Ranks
3.9	4.5	4.8	13.5
4.3	9	3.6	1
4.7	12	4.5	10
3.7	2	3.9	4.5
4.2	8	4.6	11
4.1	7	4.0	6
4.8	13.5	3.8	3
5.3	17		
4.9	15		
5.2	16		
5.5	18		
Total	122 (R_1)		49 (R_2)

(d) As one of the samples has less than 8 cases, the smaller of the two computed U values, viz., $U_1 = 21$, is compared with the critical two-tail U value for the chosen α of 0.02 and the given n_1, n_2 combination (Table E of Appendix).

$$\text{Critical } U_{.02(11,7)} = 12.$$

Because the computed U_1 of 21 is higher than the critical U for the chosen significance level, there is no significant difference between the sample means ($P > 0.02$).

Example 9.6.2.

The strengths of kneejerk reflexes (degrees of arc) under relaxed condition were found to be as follows in 9 athletes and in 11 nonathletes.

Athletes : 31, 30, 22, 30, 26, 28, 19, 36, 37.
 Nonathletes: 35, 26, 14, 20, 11, 14, 21, 31, 27, 24, 10.

Is there a significant difference between the mean strengths of kneejerks of the two groups? ($\alpha = 0.05$).

Solution :

A two-tail U test for large groups is undertaken.

(a) The scores of both groups taken together are assigned ranks in an ascending order, giving average ranks to the tied scores (Table 9.26).

(b) The ranks of each group are totalled separately to give the respective sums of ranks : $R_1 = 121$; $R_2 = 89$.

(c) The statistic U is computed from any of the rank sums.

$$n_1 = 9 ; \quad n_2 = 11.$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 = 9 \times 11 + \frac{9 \times 10}{2} - 121 = 23.$$

Table 9.26. U test of kneejerk data.

Athletes		Nonathletes	
Kneejerk strengths (X_1)	Ranks	Kneejerk strengths (X_2)	Ranks
31	16.5	35	18
30	14.5	26	10.5
22	8	14	3.5
30	14.5	20	6
26	10.5	11	2
28	13	14	3.5
19	5	21	7
36	19	31	16.5
37	20	27	12
		24	9
		10	1
Total	121 (R_1)		89 (R_2)

(d) The value of U_e expected from the H_0 is then computed.

$$U_e = \frac{n_1 n_2}{2} = \frac{9 \times 11}{2} = 49.5.$$

(e) The difference ($U_1 - U_e$) is next transformed into the z score.

$$s_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{9 \times 11 (9 + 11 + 1)}{12}} = 13.16.$$

$$z = \frac{U_1 - U_e}{s_U} = \frac{23 - 49.5}{13.16} = -2.01 ; \text{ i.e., } z = 2.01.$$

(f) Ignoring the negative sign of the computed z , the two-tail probability P of the H_0 being correct is worked out using the unit normal curve areas (Table A).

$$P = 2 [0.5000 - (\text{fractional area of the unit normal curve from } \mu \text{ to the computed } z \text{ of } 2.01)].$$

$$= 2 [0.5000 - 0.4778] = 0.044.$$

$$\alpha = 0.05.$$

Because $P < \alpha$, there is a significant difference between the mean kneejerk strengths of the groups ($P < 0.05$).

Example 9.6.3.

The memory test scores of two groups of students are given in the 1st and 3rd columns of Table 9.27. Is there any significant difference between the mean memory scores of the two groups?

Solution :

As each of the groups has more than 8 cases, a two-tail U test for large groups is undertaken.

Table 9.27. U test of memory score data.

Group 1		Group 2	
Memory scores (X_1)	Ranks	Memory scores (X_2)	Ranks
20	10.5	18	7
33	16	14	5
25	12	17	6
34	17	11	1
29	14	12	3
27	13	20	10.5
39	19	19	8.5
32	15	12	3
36	18	12	3
19	8.5		
Total	143.0 (R_1)		47.0 (R_2)

(a) The scores of *both groups taken together* are assigned ranks in an ascending order, giving average ranks to the tied scores (Table 9.27). The ranks of each group are then *totalled separately* to give the respective rank sums.

$$R_1 = 143.0 ; \quad R_2 = 47.0 ; \quad n_1 = 10 ; \quad n_2 = 9.$$

(b) The statistic U is worked out using any of the rank sums. Thus, using R_2 ,

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 = 10 \times 9 + \frac{9(9 + 1)}{2} - 47 = 88.$$

(c) The value of U_e expected from the H_0 as well as the $SE(s_U)$ of U is worked out and used in computing the z score.

$$U_e = \frac{n_1 n_2}{2} = \frac{10 \times 9}{2} = 45 ; \quad s_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{10 \times 9 (10 + 9 + 1)}{12}} = 12.25 ;$$

$$z = \frac{U_2 - U_e}{s_U} = \frac{88 - 45}{12.25} = 3.51.$$

(d) The two-tail P of the H_0 being correct is worked out using the unit normal curve areas (Table A of Appendix).

$$P = 2 [0.5000 - (\text{Area of unit normal curve from its } \mu \text{ to the computed } z \text{ of } 3.51)] \\ = 2 [0.5000 - 0.4998] = 0.0004.$$

As P is found to be lower than even 0.0005, the H_0 is rejected and it is inferred that there is a *significant difference* between the mean memory scores of the two groups ($P < 0.0005$).

9.7 MEDIAN TEST

This is a nonparametric alternative to the Student's t test for the *significance of differences between unpaired observations of two or more independent groups*. It is less powerful than the t test and the Mann-Whitney U test. It can be applied to both *continuous and discontinuous variables, irrespective of the normality or non-normality of their population distributions, to groups of identical or unequal sizes, and also to small groups or samples*. But it should be justifiable to assume that each score occurs in the sample at random and independent of all other scores.

The H_0 contends that all the sets of scores have come from the same population and consequently have an *identical median*. The H_0 proposes in effect that there are equal numbers of scores in each group above and below the common median, any observed deviation from this distribution being due to mere chances of random sampling. The probability P of the correctness of this H_0 is found out by a *chi square test of independence*.

(a) A common median is first computed for the scores of all the groups (pages 43-44).

(b) In each group, positive (+) signs are given to the scores higher than the common median while negative (-) signs are assigned to those equal to or lower than the median. For each group, the frequencies of scores with positive and negative signs are counted separately and entered as the respective f_o .

values in a *contingency table*. The latter is framed with two columns representing the positive and negative deviations of scores from the median, and as many rows as the number of samples or groups — a 2×2 -fold contingency table results in case of two samples only.

(c) A chi square test of independence (pages 203-204) is then performed, computing the expected frequencies (f_e) of the deviations on the basis of the H_0 . Where f_r and f_c are the marginal totals, i.e., the totals of cell frequencies of rows (r) and columns (c) respectively, and n is the total number of scores of all the groups, the f_e of any cell of the contingency table is given by :

$$f_e = \frac{f_r f_c}{n}; \quad df = (r - 1)(c - 1);$$

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}.$$

In case of a 2×2 -fold contingency table, an alternative formula, using the f_o values of its cells, A, B, C and D, may be applied.

$$\chi^2 = \frac{n(AD - BC)^2}{(A + B)(A + C)(B + D)(C + D)}.$$

(d) The computed χ^2 is compared with critical χ^2 values for different significance levels. If the computed χ^2 either exceeds or equals the critical χ^2 for the chosen α , it is considered significant. The scores of different samples then differ significantly ($P \leq \alpha$).

Example 9.7.1.

Following scores were obtained by 14 male and 16 female students in an English-usage test.

Males : 22, 27, 29, 30, 32, 34, 39, 45, 46, 49, 49, 50, 51, 51.

Females : 22, 23, 24, 26, 29, 31, 32, 32, 33, 35, 37, 38, 40, 40, 52, 53.

Use the median test to find whether the test scores differ significantly in the two sexes.

Solution :

A two-tail median test is undertaken.

(a) A common median is computed for the scores of both the samples.

$$n = n_1 + n_2 = 14 + 16 = 30. \quad Mdn = \frac{n+1}{2} \text{th score} = \frac{30+1}{2} \text{th or } 15.5 \text{th score.}$$

Counting off the scores of both the groups taken together and in an ascending order,

$$15 \text{th score} = 34 ; \quad 16 \text{th score} = 35 ; \quad Mdn = \frac{34+35}{2} = 34.5.$$

(b) The scores of each group are assigned positive (+) and negative (-) signs according as they are respectively higher than and lower than (or equal to) the *Mdn*.

Males : (i) Scores given negative signs : 22, 27, 29, 30, 32, 34.
Total number of negative deviations : 6.

(ii) Scores given positive signs : 39, 45, 46, 49, 49, 50, 51, 51.
Total number of positive deviation : 8.

Females : (i) Scores given negative signs : 22, 23, 24, 26, 29, 31, 32, 32, 33.
Total number of negative deviations : 9.

(ii) Scores given positive signs : 35, 37, 38, 40, 40, 52, 53.
Total number of positive deviations : 7.

(c) A 2×2 -fold contingency table is framed for computing χ^2 and the numbers of two types of deviations are entered as f_o values in its respective cells (Table 9.28). The marginal totals of cell frequencies of each row and each column are entered as f_r and f_c in the table.

(d) Because it is a 2×2 -fold contingency table, χ^2 may be computed straightway from the f_o in the cells, A, B, C and D, of the table and the marginal totals f_r and f_c .

Table 9.28. 2×2 -fold contingency table for median test.

Groups	Negative deviations (f_o)	Positive deviations (f_o)	Total (f_r)
Males	6 (B)	8 (A)	14 (A + B)
Females	9 (D)	7 (C)	16 (C + D)
Total (f_c)	15 (B + D)	15 (A + C)	30 (n)

$$\chi^2 = \frac{n(AD - BC)^2}{(A+B)(A+C)(B+D)(C+D)} = \frac{30(8 \times 9 - 6 \times 7)^2}{14 \times 15 \times 15 \times 16} = 0.536.$$

$$df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1,$$

where r and c are respectively the numbers of rows and columns of the table, containing f_o values.

(e) Critical χ^2 scores are quoted from Table C of Appendix.

$$\chi^2_{01(1)} = 6.64 ; \quad \chi^2_{05(1)} = 3.84.$$

Because the computed χ^2 is lower than the critical χ^2 for even the 0.05 level, it is not significant. So, the test scores of the two sexes do not differ significantly ($P > 0.05$).

Example 9.7.2.

The strengths of patellar reflex (in ° radian) were found as follows in three groups of young men.

Group I	: 17, 19, 22, 25, 25, 30, 31, 33, 33, 34, 36, 37, 37.	($n_1 = 13$).
Group II	: 14, 16, 18, 20, 21, 21, 23, 25, 29, 30, 30, 35.	($n_2 = 12$).
Group III	: 13, 15, 17, 19, 21, 23, 24, 30, 30, 31, 31, 32.	($n_3 = 12$).

Find if there is any significant difference between the scores of the three groups.

Solution :

(a) A common median is first computed for the scores of all three groups.

$$n = n_1 + n_2 + n_3 = 13 + 12 + 12 = 37.$$

$$Mdn = \frac{n+1}{2} \text{th score} = \frac{37+1}{2} \text{th or 19th score.}$$

Counting off the scores of all the groups taken together and in an ascending order, the 19th score is seen to be the second of three identical scores, viz., 25 each. So, to reach the *Mdn*, one of these identical scores, viz., the first 25, has to be counted off. The three identical scores, forming the set of 25, are assumed to occupy one unit interval extending from 24.5, each of them occupying 1/3 or 0.33 of this interval. As one such score is counted off for reaching the median,

$$Mdn = 24.5 + 0.33 = 24.83.$$

(b) The scores of each group are assigned positive (+) and negative (-) signs according as they are respectively higher than and lower than (or equal to) the *Mdn*.

Group I : (i) Scores given negative signs : 17, 19, 22.
Total number of negative deviations : 3.

(ii) Scores given positive signs : 25, 25, 30, 31, 33, 33, 34, 36, 37, 37.
Total number of positive deviations : 10.

Group II : (i) Scores given negative signs : 14, 16, 18, 20, 21, 21, 23.
Total number of negative deviations : 7.

(ii) Scores given positive signs : 25, 29, 30, 30, 35.
Total number of positive deviations : 5.

Group III : (i) Scores given negative signs : 13, 15, 17, 19, 21, 23, 24.
Total number of negative deviations : 7.

(ii) Scores given positive signs : 30, 30, 31, 31, 32.
Total number of positive deviations : 5.

(c) A 3×2 -fold contingency table is framed for computing χ^2 and the numbers of two types of deviations are entered as f_o values in the respective cells (Table 9.29). The marginal totals of cell frequencies of each row and each column are entered as f_r and f_c in the table. The df of the χ^2 to be computed is given by : $df = (r - 1)(c - 1) = (3 - 1)(2 - 1) = 2$.

(d) That many cells as the df (viz., 2) are randomly chosen, and the expected frequency f_e of each such cell is computed using the f_r and f_c of the row and the column to which the cell belongs. For example, for the cell giving positive deviations for Group I,

$$f_r = 13 ; \quad f_c = 20 ; \quad f_e = \frac{f_r \times f_c}{n} = \frac{13 \times 20}{37} = 7.0.$$

Similarly, for the cell giving negative deviations for Group III,

$$f_r = 12 ; \quad f_c = 17 ; \quad f_e = \frac{f_r \times f_c}{n} = \frac{12 \times 17}{37} = 5.5.$$

The f_e values for the remaining cells are obtained by subtracting the already computed f_e values from either f_r or f_c values (Table 9.29).

(e) The values of f_o and f_e of each group are used in computing χ^2 .

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(3-6.0)^2}{6.0} + \frac{(10-7.0)^2}{7.0} + \frac{(7-5.5)^2}{5.5} + \frac{(5-6.5)^2}{6.5} + \frac{(7-5.5)^2}{5.5} + \frac{(5-6.5)^2}{6.5} = 4.30.$$

Table 9.29. 3×2 -fold contingency table for χ^2 test of patellar reflex data.

Groups	Negative deviations		Positive deviations		f_r
	f_o	f_e	f_o	f_e	
I	3	6.0	10	7.0	13
II	7	5.5	5	6.5	12
III	7	5.5	5	6.5	12
f_c	17	17	20	20	37 (n)

(f) Critical χ^2 scores ($df = 2$) for different significance levels are quoted from Table C of Appendix.

$$\chi^2_{.05(2)} = 5.99 ; \quad \chi^2_{.02(2)} = 7.82 ; \quad \chi^2_{.01(2)} = 9.21.$$

As the computed χ^2 is found to be lower than the critical χ^2 scores for even 0.05 level, it is considered not significant. Thus, there is *no significant difference* between the scores of different groups ($P > 0.05$).

GLOSSARY

chi square : nonparametric statistic given by the sum of the ratios of squared deviations of the observed frequencies of a distribution from the frequencies, expected from a proposed distribution, and the respective expected frequencies.

chi square test : nonparametric analysis of frequencies to find whether or not an observed frequency distribution differs significantly from a proposed frequency distribution.

chi square test for goodness of fit : nonparametric test for finding whether or not an observed frequency distribution fits significantly with an expected frequency distribution based on a proposed distribution like the normal, binomial or Mendelian distribution.

chi square test of independence : nonparametric test to find whether or not there is a significant association between two variables.

composite rank test : nonparametric test to find the significance of difference between means of two equal-size independent groups, using the sums of the ranks (of the respective groups) given in a composite manner to the scores of both groups taken together.

G test : log likelihood ratio test for nonparametric analysis of frequencies to find if an observed distribution differs significantly from a proposed distribution.

- G test for goodness of fit** : log likelihood ratio test for finding whether or not an observed distribution fits significantly with an expected distribution, based on some proposed distribution like the normal, binomial or Mendelian distribution.
- G test of independence** : log likelihood ratio test for finding whether or not there is a significant association between two variables.
- Mann-Whitney U test** : nonparametric test to find if there is a significant difference between means of two unequal independent groups, using the sums of the ranks (of the respective groups) given in a composite manner to the scores of both groups taken together.
- median test** : nonparametric test for finding the significance of difference between two or more independent groups, using a common median of those groups.
- nonparametric statistic** : a statistic worked out without using any precomputed statistic as an estimate of parameter.
- signed rank test** : nonparametric test for finding the significance of difference between means in a single-group/matched-pair group experiment, worked out by giving ranks bearing the respective algebraic signs to the differences between the paired scores of each individual or case.
- Yates' correction** : correction to be applied on each difference between the observed and the expected frequencies in a chi square test, if any expected frequency is less than 5 and the chi square moreover has the *df* of 1 only.

10. PSYCHOLOGICAL TEST CONSTRUCTION

Psychological tests are undertaken mainly to study individual differences in behavioral or psychological variables such as intelligence, memory, aptitude, ability, personality, attitude, aspiration, anxiety and emotionality.

Psychological measurements are expressed in various ways such as the speed of response, the number of correct responses, the number of trials for achieving a given performance level, and the average number of items remembered after a brief exposure. Units equidistant on a psychological scale are assumed to represent identical differences in the given psychological variable.

10.1 PSYCHOLOGICAL VARIABLES

Psychological variables include such variables, many of which cannot be observed directly from outside, can only be inferred from expressions, behaviours and verbal reports of individuals, and consequently depend for proper evaluation on the cooperation of the subjects involved. They include intelligence, memory, aptitude, ability, attitude, aspiration, anxiety, emotions, personality and motivation. Many of them are hypothetical and abstract in nature, cannot be precisely measured on quantitative scales, and can only be assessed qualitatively. Even when quantitatively measurable, some of these variables have an *interval scale* with an arbitrary zero point instead of a real zero ; however, some psychological variables, such as the ratios of psychophysical stimuli and calory expenditures in job activities in industrial psychology, are measured quantitatively in *ratio scales* with real zero points (page 3).

Psychological experiments involve variables such as dependent, independent, extraneous, relevant and intervening variables.

Dependent variables

This is the behavioral response to be measured or studied in an experiment after exposure of the subjects to different levels of the independent variable(s), for assessing the effects of the latter on that behavioral response. Dependent variable can be measured by the number of correct responses to a stimulus, the time taken to react to the given stimulus, and the accuracy of performance. Sometimes, the dependent variable is measured by objective tests using rating scales. Thus, the dependent variable is often the measured behavioral response to the given independent variable(s), which constitutes mostly quantitative and continuous data, but sometimes qualitative data as in personality tests of projective type.

Independent variables

Independent variables are *deliberately chosen* by investigators and used in experiments for studying their effects on specific dependent variables. In psychological experiments, they belong to two types.

(a) *Organismic variables* include (i) *physical characteristics* of the subjects such as their sex, age, eye color, body build, height and weight, any of which may be chosen as the independent variable by the investigator, and (ii) *psychological characteristics* of the subjects such as their intelligence, personality factors, drive, emotionality, neuroticism, extroversion, aspiration level, motivation, anxiety, tension and frustration. In correlational research, the independent variables used are mainly such physical or psychological characteristics of the subjects. Such independent variables can rarely be manipulated directly (or "fixed") by the investigator ; for example, the latter cannot directly manipulate the intelligence, personality,

age or sex of a subject. So, such independent variables are liable to random changes and may be considered as *classification variables* (page 5). Nevertheless, such a variable can be manipulated *indirectly* through a *selection procedure* like the choice of subjects with specific required levels of intelligence.

(b) *Stimulus variables* consist of such environmental events including both physical and social variables, which stimulate specific receptors of the subjects to affect the dependent variable, viz., a specific behaviour of the subjects. The investigator can *directly* manipulate (or "fix") the stimulus variable chosen as the independent variable, such as changes in the intensity of the stimulating light, in the number of syllables offered in memory experiments, in the color of the light stimulus in an experiment on after-images, in the pitch of a sound stimulus or in the decibels of noise used as the independent variable in experiments on attention, or in the instructions for reaction time experiments. Such stimulus variables, being under the manipulative control of the investigator, are not liable to random changes and may be considered as "fixed" treatment variables (page 5).

Extraneous variables

These are numerous variables which occur or arise in the physical or social environment, in the subjects under study, or in the experimental procedure, but are not intended to be used as the dependent and independent variables in the experiment being undertaken. Extraneous variables include such physical and social environmental factors as well as such physical and psychological characteristics of the subjects, as are other than the dependent and independent variables. They are additional variables happening to occur in any experiment, such as noise, temperature, light, drought and humidity in the laboratory, the order of presenting the stimuli in a reaction time

experiment, and the size of printed materials used in a memory experiment, none of which is chosen or deliberately used by the investigator for the purposes of the experiment. According as the dependent variable being studied may or may not be affected by such extraneous variables, they belong to two classes.

(a) *Relevant variables* : These are such extraneous variables which, though not deliberately used or intended to be used by the investigator to study their effects on the dependent variable in that particular experiment, occur spontaneously, can influence and affect the dependent variable, and may consequently defeat the purpose of the experiment to study the effect of only the independent variable on the dependent one. The investigator must remain vigilant about these relevant variables and must control them as far as possible so as to minimize their effects on the dependent variable. They can be sought to be controlled by methods like constancy of experimental conditions, balancing and counterbalancing, randomization, matching, and changes in design in multiple-group experiments. Relevant variables are further classified into subject relevant, situational relevant and sequence relevant variables. (i) *Subject relevant variables* are *organismic variables* owing to both *physical characteristics* (e.g., age, sex, race, etc.) and *psychological characteristics* (e.g., intelligence, neuroticism, personality factors and motivational aspects) of the subjects under study. It is very hard to control these organismic variables because of the difficulties in assessing them from outside and in manipulating them directly. (ii) *Situational relevant variables* are those which occur in the experimental situation and the environment ; e.g., light, noise, temperature and humidity in the laboratory, distractions, etc. They can mostly be controlled by the

investigator. (iii) *Sequence relevant variables* arise from the sequence of applications of the independent variable and include fatigue, practice, monotony, etc. (See page 6 also.)

(b) *Irrelevant variables* : These are such variables which do not perceivably affect the dependent variable. For example, hair color, eye color, skin complexion or economic condition of the subjects may be considered irrelevant variables in an experiment to study the effect of practice on memory.

Intervening variables

It is difficult to identify and control some such psychological organismic variables as act side by side with the independent variable, without the investigator being aware of that, and affect the dependent variable. In an instrumental conditioning experiment, drive may be such a variable, and it may link an independent variable like food deprivation with a dependent variable such as the behavioral modification. In industrial work, lack of motivation may be such a variable and the real cause of decrement of the output rather than physiological fatigue considered as the independent variable affecting the dependent variable of industrial production. In an experiment studying the effect of intelligence level on achievement, anxiety may be such a variable to influence achievement (dependent variable) side by side with intelligence (independent variable), and consequently needs to be neutralized first. Such unobserved hypothetical variables, assumed to be associated with independent variables like intelligence, emotion, motivation, aspiration and habit, are based on logical constructs and are called *intervening variables*.

10.2 PRINCIPLES OF TEST CONSTRUCTION

This chapter deals mainly with the tests of abilities. Essentials of test construction are

outlined below.

(a) *Area to be assessed* : First, the specific area of ability to be assessed should be identified.

(b) *Selection of test items* : A number of test items should be so chosen for each test as to enable the proper and efficient exploration of the intended area of ability. A dichotomously scored test item is scored as +1 or 0 according to right or wrong answers respectively, and the total of item scores gives the test score.

(c) *Item analysis* : As items are selected depending on their *difficult values* and *discriminatory powers*, item analysis should be undertaken to determine these two properties of each test item (§ 10.4).

(d) *Arrangement of test items* : The test items selected by item analysis should be serially arranged in a suitable order according to the type of test (§ 10.3).

(e) *Reliability* : Reliability of a test is the consistency of results on its repeated applications on the same sample under identical conditions. Before putting a test to actual use, its reliability should be estimated by applying it on a properly drawn representative sample (§ 10.5).

(f) *Validity* : Validity is the capacity of a test to measure a specific variable in exclusion of others. It should next be estimated from the scores of the test in a representative sample (§ 10.6). *Factor analysis* should be undertaken for the construct validity of a test (§ 10.8).

(g) *Standardization* : Methods of administration and scoring should be precisely laid down for the test and standardized (§ 10.9).

(h) *Establishment of norms* : A norm is the average score of a representative group of subjects in a test in terms of a convenient scale of transformed scores. For interpreting the test

scores, the scores of a representative group should be statistically treated to establish the norms (§ 10.9).

10.3 POWER AND SPEED TESTS

Tests belong to two types, power tests and speed tests, according as they differentiate between individuals in terms of difficulty levels of items and the speed of performance, respectively.

In *power tests*, the test items are usually grouped according to the types of contents and are arranged serially in an increasing order of difficulty. Enough time is given to enable at least 75% of the subjects being tested to attempt all the items. Still, some may fail to answer all the test items because of the difficulty level. Achievement tests are pure power tests.

In *speed tests*, all the test items are of uniform difficulty level, and the subject being tested has to answer the test items within a stipulated time which is so short that none can answer all the items in time. The level of ability of the subject is determined by the latter's speed of performance.

In most tests, however, both speed and power are mixed in varying degrees. So, there is no rigid line of demarcation. Estimation of reliability depends on the proportions of power and speed in the test.

10.4 ITEM ANALYSIS

Most tests are composed of test items. This chapter is mostly concerned with test items of ability tests. Item analysis is essential for selecting suitable items which confirm and enhance the reliability and validity of the test and contribute towards the goal of the test. Properties like score distribution, mean and variance of the total test scores depend on the properties of the item ; so the need for item analysis.

Item analysis determines the difficulty level and the discriminatory power of each test item — the discriminatory power of the test is, in a broader sense, an index of validity. For such determination,

(a) the psychological processes, involved in the attribute under study, are first analyzed ;

(b) a list of items is next prepared with more items than what is required for the given test ;

(c) these items are then administered to a representative sample on a trial basis ;

(d) the selection of test items is finally done quantitatively by determining for each item : (i) the *difficulty value* given by the percentage or proportion of individuals answering a test item correctly, and (ii) the *discriminatory power* which is the ability of a test item to discriminate between individuals with respect to the given variable ;

(e) it should be ensured that the test items should contribute equally in predicting the psychological variable.

Item analysis is done for the final selection of items for a test closer to power test ; but it is useless for items meant for a pure speed test — for the latter, the items are instead rotated in different orders so that every item has an equal chance of going to the last part of the test and the limited time allotted is balanced.

For ability tests, items are selected depending on the difficulty value of dichotomously scored items. For non-ability tests, items should better be selected by considering the item-total correlation as well as the difficulty value.

1. Difficulty value

This is determined for a test item by the proportion p of right answers to that item by the individuals of a sample. Where n is the total number of individuals in the sample, P is

the number of individuals passing or giving the right answer to the given test item, and p and q are the proportions of individuals of the sample giving respectively right and wrong answers to the item,

$$p = \frac{P}{n}; \quad q = 1 - p.$$

Some important properties of the difficulty value are discussed below.

(a) The difficulty level of any item is actually inversely related to its p value. The latter (p) represents the average item score as well as the mean index of difficulty for the individuals. For individuals of the same population, items with an identical p value are considered as equally difficult.

(b) Because a dichotomous item is scored as 1 or 0, p is also the mean score of all candidates in that item and is given by dividing the total scores of both successful and unsuccessful candidates in that item by the sample size n . The greater is this mean score p of the item, the lower its difficulty value and discriminatory power. Thus, p is a direct measure of the easiness of an item and an indirect measure of its difficulty. If, 60 out of 150 persons answer correctly the item 2 of a test of 7 items, the p value of this item is given by $60/150$ or 0.40 which is the proportion of "pass" in this item. This item would be considered less difficult than another with the p value of 0.30, but more difficult than one with the p value of 0.70.

(c) The variance (s^2) and the standard deviation (s) of the scores in an item depend on its p value.

$$s^2 = pq; \quad s = \sqrt{pq}.$$

Thus, the p value affects the nature and shape of the score distribution of the item.

It is desirable to choose an item with p value of 0.50 so that its scores have a variance

of 0.25.

$$p = 0.50;$$

$$q = 1 - p = 1 - 0.50 = 0.50;$$

$$s^2 = pq = 0.50 \times 0.50 = 0.25.$$

Such p value provides information about each individual passed or failed. In contrast, an item with p value of 1 or 0 gives little information about individual differences. However, it is difficult to discriminate individuals with different scores if all the items possess an identical p value of 0.50, because the items of a test, expected to reveal the same attribute, would be correlated with each other. It is, therefore, preferable to select test items having an average p value of 0.50, but with p values of individual items moderately spread out around that average.

(d) The p value is related indirectly to the reliability of a test, because a high p value is associated with a high item-to-item correlation.

(e) For dichotomously scored multiple choice items, the p value is affected by both item difficulty and guessing or chance factor. The chance factor would enable some individuals to hit upon the right answer to a test item by guesswork. Guessing would thus cause the obtained proportion p of passing an item to exceed the real proportion and the item will consequently appear to be relatively easy. The obtained p of right answers to the item should, therefore, be corrected for the guess factor by Guilford's formula to give the correct proportion p_c , using the number k of alternative choices for answer to the item :

$$p_c = \frac{kp - 1}{k - 1}.$$

(f) Rise in the number of alternative answers to a multiple-choice test item would increase the difficulty level of the item owing to the decline in the chance factor. So, to avoid the lowering of reliability owing to the guess factor, the desired uncorrected p should be set

lower for an item affected by guess factor than that for an item free from the guess factor, but progressively higher with the rise in the number of alternative answers. Thus, where p may be set at about 0.69 for a 5-alternative item, it should be set at about 0.67 for a 4-alternative one.

(g) Difficulty levels, determined by p values, have no linear relationship with ability. The p values simply show the relative values of items and do not express the absolute difference between two items, because units of p are not equal-interval ones. So, equal differences in the proportions (p) do not express an identical difference in difficulty. For this, the p values should be transformed into z scores using the unit normal curve area. This produces a linear scale with increasing order of difficulty. It also provides a correction for chance factor in multiple-choice items.

If, for example, 25%, 32% and 40% of a group of students were successful in answering test items A, B and C respectively, it indicates that A is more difficult than B while C is the easiest of the three. Assuming a normal distribution of the ability measured by the test, the relative difficulties of A, B and C can be converted to σ values and shown in a normal distribution with equal-interval units. Thus, 25% or 0.2500 of the unit normal curve area in its right or high-value tail represents the proportion (p) of students successful in A. Fractional area of the unit normal curve from its centre to the lower limit of this area for success in A is given by the difference between half the normal curve area and the area for success in the item : $0.5000 - 0.2500$; referring this area to the unit normal curve table, it is seen that this area corresponds to the interval $\mu + 0.6745\sigma$ (Table 6.2 and Fig. 6.2 (c) and Fig. 10.1). Thus, 0.675σ is the lower limit of the area for success in A and gives the σ value of difficulty for A (Table 10.1). Similarly, 32% or 0.3200 of the normal curve

Table 10.1. σ values of difficulty of test items.

Test items	% success	σ value of difficulty	σ difference
A	25	0.675σ	$A - B = 0.207\sigma$
B	32	0.468σ	
C	40	0.253σ	$B - C = 0.215\sigma$

area in its right tail gives the proportion (p) of students successful in B. The fractional area of the unit normal curve from its centre to the lower limit of this area for success in B amounts to : $0.5000 - 0.3200 = 0.1800$, which corresponds to the interval $\mu + 0.468\sigma$ (Table A of Appendix). Thus, 0.468σ is the lower limit of success in B and gives the σ value of difficulty for B (Table 10.1). Again, 40% or 0.4000 of the unit normal curve area in its right tail represents the proportion (p) of students successful in item C. The fractional area of the unit normal curve from its centre to the lower limit of this area for success in C amounts to : $0.5000 - 0.4000 = 0.1000$, which corresponds to the interval $\mu + 0.253\sigma$ (Table A). Thus, 0.253σ is the lower limit of success in C and gives the σ value of difficulty for C (Table 10.1).

It is seen from the above example that the higher the p value of an item, the lower is the z score for the lower limit of the corresponding

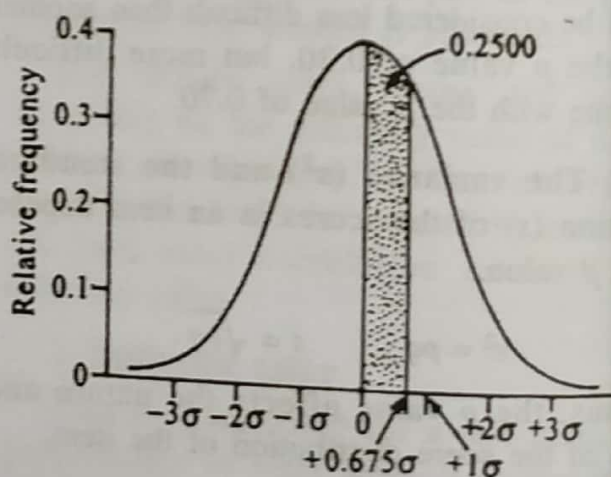


Fig. 10.1. Fractional area of the interval $\mu + 0.6745\sigma$ of the unit normal curve.

fractional area in the right tail of the unit normal curve. In the example cited above, the test.

item A : $p = 0.25$; $z = 0.675\sigma$;

item B : $p = 0.32$; $z = 0.468\sigma$;

item C : $p = 0.40$; $z = 0.253\sigma$.

Thus, where $p = 0.50$, $z = 0.00\sigma$, i.e., the μ of the unit normal curve. It, therefore, follows that the proportions (p) of successful candidates, having negative and positive z scores, amount respectively to >0.50 and <0.50 .

(h) It is possible to calculate the proportion of students expected to be successful in test item D, if D is more difficult than A by twice the relative difference in difficulty between B and C in terms of their σ difference.

σ value of difficulty for D

$= (\sigma \text{ value for A}) + 2 \times (\sigma \text{ difference of B - C})$

$= 0.675\sigma + 2 \times 0.215\sigma$

$= 1.105\sigma$.

So, $+1.105\sigma$ marks the lower limit for the unit normal curve area corresponding to the proportion p of students successful in item D. Fractional area of the unit normal curve from its centre to $+1.105\sigma$ amounts to 0.3654 (Table A of Appendix). So, the remaining area in the right tail beyond 1.105σ amounts to : $0.5000 - 0.3654 = 0.1346$. This area gives the probability of cases passing in the item D. Thus, 13.5% of the candidates are expected to pass in item D.

2. Discriminatory value

The discriminatory value of an item indicates its ability to discriminate between the subjects of a test, belonging to upper and lower categories with respect to their answers to that item. The discriminatory value of a test may be estimated by (i) *item validity* measured by item-criterion correlation, (ii) *item-total correlation* between the scores in that item and total scores of the test, and (iii) *index of discrimination*, according to the purpose of

(1) *Item validity or item-criterion correlation* :

This is determined by correlating the scores of a given test item with a criterion. A high correlation with the criterion indicates a high discriminatory value of that item. The criterion to be chosen depends on the type of validity envisaged for the test — content, construct or predictive validity. When emphasis is laid on construct validity, the criterion used is the total score of the same test or of any other equivalent test. Measures of job performance, teacher's rating or grade performance maybe used as criterion for content validity. For tests of interest and personality, validation is done by correlating the items with a suitable external criterion revealing the given attribute. For achievement tests, the criterion used may consist of grade performances or teacher's ratings. Academic achievement is often used as the criterion for intelligence test. When emphasis is laid on predictive validity, such an external criterion is used as consists of the measures of everyday success in the variable under investigation.

(2) *Internal consistency or item-total correlation* :

The test items may sometimes be correlated with the total scores of that test for internal consistency (*item-total correlation*). The homogeneity of a test depends upon the highest average item-total correlation.

According to the type of validity required and the nature of variable, point biserial r , biserial r , tetrachoric r or phi coefficient may be computed to correlate the scores in the item with either the total test scores or the criterion revealing the desired attribute.

(a) *Point biserial r or r_{pbi}* is used in correlating the scores of a continuous variable like the total test scores with a genuinely

dichotomous variable such as the right/wrong, yes/no and 1/0 answers to a test item (§ 8.7). Thus, it may be used for *item-total correlation* for item analysis (Example 8.7.3). r_{pbi} is also preferred for correlating a dichotomously answered test item with an external criterion revealing the attribute under investigation (*item-criterion correlation*), if the predictive power of the item for the given attribute is to be explored.

(b) *Biserial r* or r_b is used in correlating the scores of a continuous variable like the total test scores with the answers to an *apparently* dichotomous test item. In such cases, r_b is applied for item-total correlation (§ 8.8 and Example 8.8.1). Such *item-total correlation* measured by r_b is independent of the difficulty value of the test item. r_b is also computed between a dichotomized test item and the continuous scores of an external criterion in order to assess whether the item measures the same attribute as represented by the criterion (*item-criterion correlation*).

(c) *Phi coefficient* (ϕ) is used for *item-to-item correlation* between two dichotomous test items, each scored as yes/no, right/wrong or 1/0 (§ 8.9). It is also used in correlating a dichotomously scored test item and a genuinely dichotomous external criterion such as success/failure in a public examination (*item-criterion correlation*) so as to explore the power of the test item to predict one of the two categories of the dichotomous attribute represented by the criterion.

Phi coefficient may be computed directly from the observed frequencies (f_o) of cases in different combinations of the classes of two dichotomous variables (either two dichotomous items or an item and a dichotomous criterion), arranging these classes along the rows and columns of a 2×2 -fold contingency table (Example 8.9.2). However, it is often computed also from the proportions of cases in different combinations of the classes of two dichotomous

variables being correlated — for example, the proportions of cases passing or failing in one or both test items being correlated (Example 10.4.1). In the latter case, where a , b , c and d are the proportions of cases in the four cells of the 2×2 -fold contingency table, p_1 and p_2 are the proportions of passes or right answers in items 1 and 2 respectively, and q_1 and q_2 are respectively the proportions of failures or wrong answers in those items (Table 10.3),

$$\begin{aligned}\phi &= \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \\ &= \frac{ad - bc}{\sqrt{p_1 p_2 q_1 q_2}} ; \\ \text{or, } \phi &= \frac{a - p_1 p_2}{\sqrt{p_1 p_2 q_1 q_2}}.\end{aligned}$$

However, when an item-criterion correlation is worked out with a criterion dichotomized at its median into upper and lower groups, each with a proportion of 0.50 of the sample size ($p = q = 0.50$), and p_u and p_l are the proportions of pass in the respective groups,

$$\phi = \frac{p_u - p_l}{2\sqrt{pq}}.$$

(d) *Tetrachoric r* or r_t is computed for *item criterion correlation* to measure the extent to which a dichotomized test item and a dichotomized criterion estimate the same attribute, when the test item and/or the attribute may be considered artificially dichotomized (§ 8.10). It is also used in correlating the items of two questionnaires scored on scales other than dichotomous responses. It may also be applied to personality tests with dichotomized test items. However, because of its higher *SE* and lower reliability than product-moment r and a complex procedure for its significance test, r_t is seldom used in item analysis.

(3) *Index of determination* (D) :

This may also be used in determining the discriminatory value of an item. D expresses

the difference between the proportions ($p_u - p_l$) of candidates of the upper and lower groups of the dichotomized criterion passing in the given test item. D ranges from -1.00 to $+1.00$ in value. It is mainly used in item analysis of test to be conducted with small group such as a test administered to a class by a teacher. To find the difference ($U - L$) between percentages of the upper and lower criterion groups passing in the test item, usually 27% of individuals are selected from both these groups in case of

large normally distributed samples. The ($U - L$) difference between the proportions of these two selected batches passing in a test item gives the index of discrimination of that item, indicating its discriminatory power. The index of discrimination should preferably be used for test items with the difficulty value of 0.50.

(4) Factor analysis :

Some psychologists prefer *factor analysis* for assessing the discriminatory value (§ 10.8).

Example 10.4.1.

In a psychological test, 35 persons gave right answers to each of two test items, 25 persons answered both wrongly, 20 persons gave right answers to item 2 and wrong answers to item 1, and 10 persons answered item 2 wrongly but item 1 correctly. Is there any item-to-item correlation between the two items?

Solution :

(a) A 2×2 -fold contingency table is framed (Table 10.2) with its top and bottom rows for respectively right and wrong answers of one item, and its right and left columns for respectively those of the other item. The frequencies of persons, giving specific combinations of right and wrong answers to the two items, are entered in the respective cells for such combinations. For example, the frequency of persons answering item 1 rightly and item 2 wrongly, is entered in the top left cell or cell B. The marginal totals, f_r and f_c , are worked out for the rows and columns respectively.

Table 10.2. Fourfold contingency table for right and wrong answers to two test items.

Item 1	Item 2		Total f_r
	wrong	right	
right	10 (B)	35 (A)	45 (A + B)
wrong	25 (D)	20 (C)	45 (C + D)
Total (f_c)	35 (B + D)	55 (A + C)	90 (n)

(b) All the frequencies are converted to the respective proportions by dividing them with the sample size ($n = 90$) and arranged in Table 10.3.

(c) Phi coefficient is computed using the cell proportions (a, b, c and d) and the marginal proportions (p_1, p_2, q_1 and q_2) of Table 10.3.

$$\phi = \frac{ad - bc}{\sqrt{p_1 p_2 q_1 q_2}} = \frac{0.389 \times 0.278 - 0.111 \times 0.222}{\sqrt{0.500 \times 0.611 \times 0.500 \times 0.389}} = + 0.34.$$

[Alternatively,

$$\phi = \frac{a - p_1 p_2}{\sqrt{p_1 p_2 q_1 q_2}} = \frac{0.389 - 0.500 \times 0.611}{\sqrt{0.500 \times 0.611 \times 0.500 \times 0.389}} = + 0.34.]$$

(d) The computed ϕ is converted to χ^2 and compared with critical χ^2 scores.

$$\chi^2 = n\phi^2 = 90 \times (0.34)^2 = 10.40 ; \quad df = 1.$$

$$\text{Critical } \chi^2 : \quad \chi^2_{.001(1)} = 10.83 ; \quad \chi^2_{.01(1)} = 6.64. \quad (\text{Table C})$$

Because the computed χ^2 is higher than the critical χ^2 for 0.01 level, it is *significant* and there is a significant item-to-item correlation ($P < 0.01$).

Table 10.3. Fourfold contingency table for phi coefficient from proportions.

Item 1	Item 2		Total
	wrong	right	
right	0.111 (b)	0.389 (a)	0.500 (p_1)
wrong	0.278 (d)	0.222 (c)	0.500 (q_1)
Total	0.389 (q_2)	0.611 (p_2)	1.000

Example 10.4.2.

3 items were chosen out of 50 items of a psychological test administered to 200 students. The upper and lower 27% of the students, scoring the highest and the lowest scores respectively in the test, were chosen to form respectively the U and L groups, each consisting of 54 students. The number of students of each group, passing in each of the chosen items, is entered in Table 10.4. Estimate the discriminatory powers of the test items.

Solution :

(a) Proportion of each group, passing in every item, is computed by dividing the number of passed students (P) by the group size (n). For example, for the L group in case of test item 3,

$$P = 21 ; \quad n = 54 ; \quad L = \frac{21}{54} = 0.39.$$

(b) The index of discrimination (D) is computed for each test item as the difference between the proportions of pass in the upper and the lower groups (U and L). For example, for item 1, the proportions of pass in the two groups amount to :

$$U = 0.74 ; \quad L = 0.26. \quad \therefore D = U - L = 0.74 - 0.26 = 0.48.$$

Table 10.4. Computation of the index of discrimination of test items.

Test items	Number of pass		Proportion of pass		D = $U - L$
	U group (54)	L group (54)	U	L	
1	40	14	0.74	0.26	0.48
2	27	27	0.50	0.50	0.00
3	33	21	0.61	0.39	0.22

Thus, it may be inferred from the respective D scores that item 1 has the highest discriminatory power and item 2 has no power of discrimination.

10.5 RELIABILITY OF A TEST

Reliability is the consistency of results of a test when it is repeated on the same individual or group on different occasions under identical testing situations. It is also indicated by the close similarity of the scores of a test with those of other equivalent forms of the test administered to the same group. Reliability indicates the stability of scores which tend to be unstable because of errors of measurement.

Errors of measurement

Errors of measurement may be either random or systematic.

(a) Random errors :

Reliability is affected mostly by random errors. Random errors result from uncontrolled and uncorrelated factors such as : errors of scoring and interpretation ; fluctuating mood, motivation, attention or interest of the subject ; noise and other distractions. Sampling of test items may be a major source of random errors. It may cause item-to-item variations within a test. The time gap between administrations of alternate forms of a test as well as differences in their contents and scoring may result in random errors. In multiple-choice tests, guessing definitely serves as a source of random errors due to variations in item-to-item performance, and lowers the reliability.

Random errors of measurement should be minimized by proper item sampling, large sample size, clear instructions for proper test administration, and objective scoring and interpretation.

Random errors may be positive as well as negative — not systematically unidirectional for all the scores. The mean random error will be zero if a large number of trials is undertaken. Random errors are assumed to be uncorrelated with each other and with the true scores.

Variations in the scores of an individual, on

repetition of the same test, may result from either genuine changes in his ability or random errors of measurement. The *true score* (X_{∞}) is the expected average of the scores of an individual in an infinitely large number of repetitions of the test under identical conditions, and is assumed to remain unchanged because it is supposed to be free from any *random error* (X_e). The *obtained score* (X_t) of an individual differs at random from X_{∞} due to X_e ; for reliability, X_t should correspond closely to X_{∞} . X_{∞} is almost always unknown — it is estimated in practice in terms of a confidence interval around X_t . In case of a large number of observations, X_t scores should be normally distributed around X_{∞} because X_e is also expected to be normally distributed.

$$X_t = X_{\infty} + X_e ; \quad \text{or, } X_{\infty} = X_t - X_e ;$$

$$\text{or, } X_e = X_t - X_{\infty}.$$

Reliability depends on the closeness between X_t and X_{∞} with only a minimal X_e .

(b) Systematic errors :

Systematic or constant errors affect mainly the validity of a test, not its reliability. They result from defects in test construction, limited time allotment to an ability test, some personal modes of response and environmental variations. Systematic errors act unidirectionally, causing either overestimation or underestimation of the scores in all cases systematically.

Reliability coefficient

Reliability may be statistically defined as the proportion of *true variance* (s_{∞}^2) in the total variance (s_t^2) of the scores of a test. In that context, *reliability coefficient* (r_{tt}) is a measure of correlation between s_{∞}^2 and s_t^2 of the test scores. The total variance (s_t^2) of scores of a test estimates the deviations of individual scores (X_t) from their mean (\bar{X}_t). It is the mean squared deviation of X_t scores from \bar{X}_t .

$$s_t^2 = \frac{\sum (X_i - \bar{X}_t)^2}{df}$$

A part of s_t^2 is due to the true variance (s_∞^2) while its other part comes from the error variance (s_e^2). The latter may arise from errors of measurement and is not the same as sampling errors. In absence of any correlation between s_∞^2 and s_e^2 ,

$$\frac{s_\infty^2}{s_t^2} + \frac{s_e^2}{s_t^2} = 1.00.$$

The lower is the proportion of s_e^2 in s_t^2 , the closer is the true variance s_∞^2 to the total variance and consequently, the closer are X_∞ and X_t , and the higher the reliability of that test. Indeed, r_{tt} is the proportion of s_∞^2 in s_t^2 and can be considered as a self-correlation of a test.

$$r_{tt} = \frac{s_\infty^2}{s_t^2} = 1 - \frac{s_e^2}{s_t^2}; \quad \text{or, } s_e^2 = s_t^2 (1 - r_{tt}).$$

Thus, an r_{tt} of 0.70 indicates that s_∞^2 and s_e^2 constitute respectively 0.70 and 0.30 proportions of s_t^2 .

r_{tt} suffers from errors of measurement. It ranges from -1.00 to +1.00, depending on the type and objective of the test as well as the nature of the group tested. It may lie in the

Example 10.5.1.

For a psychological test, r_{tt} and s_t were found to be 0.75 and 10 respectively. Compute s_∞^2 and s_e^2 , their proportional contribution to s_t^2 , and the SE of measurement of the test.

Solution :

$$r_{tt} = 0.75; \quad s_t = 10. \quad \therefore s_t^2 = 10^2 = 100.$$

$$s_e^2 = s_t^2 (1 - r_{tt}) = 100 (1 - 0.75) = 25.$$

$$\therefore s_e = \sqrt{s_e^2} = \sqrt{25} = 5.$$

$$s_\infty^2 = r_{tt} s_t^2 = 0.75 \times 100 = 75.$$

$$\frac{s_e^2}{s_t^2} = \frac{25}{100} = 0.25; \quad \frac{s_\infty^2}{s_t^2} = \frac{75}{100} = 0.75.$$

0.50–0.60 range for a group within a narrow range of school grades; but r_{tt} should be within the 0.90–0.95 range for individual diagnosis and classification. Its magnitude and algebraic sign are interpreted like those of Pearson's r .

Reliability depends on the close agreement between the obtained score (X_t) and the true score X_∞ . The *standard error of measurement* (s_e) is the square root of the error variance (s_e^2) and is thus another measure of random errors.

$$s_e^2 = s_t^2 (1 - r_{tt}); \quad s_e = \sqrt{s_t^2 (1 - r_{tt})}.$$

When s_e is zero or negligible, the test is perfectly reliable ($r_{tt} = 1$) and X_t coincides with X_∞ ; differences between X_t scores are then solely due to genuine differences between X_∞ scores. But when s_e equals the square root (s_t) of the total variance, the test has no reliability ($r_{tt} = 0$).

Whereas r_{tt} is a measure of *self-correlation* of a test with itself, the *index of reliability* ($r_{t\infty}$) is a measure of correlation between the whole obtained score X_t and its part (X_∞) which is the true score. Thus, $r_{t\infty}$ is a *part-whole correlation*: $r_{t\infty} = \sqrt{r_{tt}}$. r_{tt} is a better measure for comparing the reliabilities of different test scores.

Estimation of reliability

Three measures of reliability, viz., coefficient of stability, coefficient of equivalence and coefficient of internal consistency, have been recognized by the American Psychological Association. These are *reliability coefficients* (r_{tt}) and are estimated respectively by the test-retest method, the alternate-forms or parallel-forms method, and the split-half method. The methods differ with respect to the contributions of s_w^2 and s_e^2 in their computation. The basic relations are given by the following formulae :

$$r_{tt} = \frac{s_w^2}{s_t^2} ; \quad \text{or, } r_{tt} = 1 - \frac{s_e^2}{s_t^2}.$$

1. Test-retest method :

In this method, the same test is repeated on the same group of individuals after a suitable interval and the two sets of scores are correlated for computing the *coefficient of stability* or the *coefficient of retest reliability* — the coefficient of stability determines the dependability of the measurement over a time interval. The error variance (s_e^2) measured by this coefficient is an estimate of random variations of scores over a time interval due to either uncontrolled testing conditions such as change of sets and weather, poor visibility or audibility, noise and distractions, or personal factors such as lack of motivation, guess work, fatigue, boredom, anxiety and awareness about earlier mistakes. Maturity factors during the time gap may also contribute to s_e^2 . Closeness of the two sets of scores indicate a low s_e^2 and a high stability coefficient (r_{tt}).

In this method, an optimum time interval should be chosen for retesting. A short interval may apparently increase s_w^2 to enhance r_{tt} owing to the memory effect. On the contrary, if the interval is too long, differential rates of physical and emotional changes of the testee may affect the scores, raising s_e^2 and lowering r_{tt} ; the interval should not exceed six months.

r_{tt} may also be enhanced due to the unchanged content which raises s_w^2 .

The test-retest method can be used for tests of psychomotor abilities, sensory discriminations and such cognitive skills as are less affected by practice and memory effects. It is suitable for speed tests. It can also be applied to heterogeneous tests where items measure different abilities and have high correlations with a criterion, but a low correlation with other items of test, making the reliability of internal consistency insignificant.

Various types of reliabilities of ratings such as performance evaluation and personal assessment ratings can be estimated with the help of the coefficient of stability. This can be done by working out (i) *inter-rater reliability*, i.e., the consistency of the results when the same individual is rated by two or more raters, and (ii) *rate-rater reliability*, i.e., the consistency of the results when the same individual is rated repeatedly by the same rater over a time gap.

The coefficient of stability (r_{tt}) gives no indication about the internal consistency of the test. The test-retest method is considered unsuitable for many psychological tests, and is used in such cases mainly in the absence of an alternate form of the test.

2. Alternate-forms or parallel-forms method :

In this method, two parallel or equivalent forms of a test are administered to the same group of individuals and the scores of the two tests are correlated for working out the *coefficient of equivalence* which gives the reliability of the original test. Equivalent or parallel forms of a test should possess an identical true variance (s_w^2), similar item difficulties, similar item-total correlations, and independent error variances (s_e^2) with no overlap. *Criteria of parallelism* for such alternate forms of a test include identical score

distributions, identical mean scores, identical variances, perfectly matched contents, equal item-total correlations, same item-intercorrelations, identical method of administration, and the same number, types and difficulty levels of test items.

The equivalent-forms reliability may be expressed in two ways.

(a) *Coefficient of stability and equivalence* :

It is the correlation coefficient between two sets of scores obtained by administering two equivalent or parallel forms of a test to the same group on two different occasions separated by a time interval. It measures both *equivalence of contents* of the two tests and *temporal stability*, i.e., the dependability of measurements over a time interval. The error variance (s_e^2) results here from both temporal variations of performance and variations of scores due to different sets of items in the two tests — the latter may be termed *item specificity*.

(b) *Coefficient of equivalence* : It is the correlation coefficient between the scores of two equivalent tests administered simultaneously or in immediate succession to the same group. It measures *only the equivalence* of the two tests and not the temporal stability. Here, s_e^2 results from variations of performance in different sets of items, or item specificity alone.

This alternate or equivalent-forms method is highly suitable for speed tests, but cannot serve for heterogeneous tests unless item-intercorrelations have been taken care of while developing the parallel tests. In practice, it is difficult to get precisely equivalent forms fulfilling all the criteria of parallelism ; the equivalent-forms reliability may be affected by any deviation from precise equivalence such as an overlap of error variances (s_e^2), inequality of either s_e^2 or true variance (s_w^2), variations of s_e^2 owing to change in content of the alternate

form, rise of s_e^2 because of different item difficulties of the parallel forms of the test, or fluctuations in testing environment and method. The time interval between the administrations of the two tests, individual changes in motivation, practice effects resulting from the use of similar item contents in the parallel forms, distractions, fatigue and boredom influence the error variance (s_e^2) to affect the equivalent-forms reliability just like the test-retest reliability. But unlike the test-retest method, the item contents are not identical — only similar — on the two occasions of administration of the parallel tests. This decreases the memory effect and the practice effect resulting from prior use. So, the alternate-forms method is frequently preferred to the test-retest method and used in a larger number of cases than the latter.

3. *Split-half method* :

In this method, a single form of test is administered to a group of individuals, but either the testing procedure or the scores is/are divided into two equivalent halves and the obtained scores of the two halves are correlated to give the *split-half reliability coefficient* (r_{hh}). This test is based on the assumption that different items of a test estimating any attribute should measure the same variable and should thus be internally consistent. So, the correlation coefficient between the scores of the half-tests acts as a measure of equivalence and consistency of the two half-tests.

The entire test may be split into two parallel halves possessing items of equivalent difficulties and identical item correlations. The two halves are then administered as separate tests to the same group either simultaneously or in immediate succession. However, because reliability is a function of test length and increases with the latter so long as its homogeneity is maintained, the split-half correlation coefficient (r_{hh}) is extended by the *Spearman-Brown formula* to give the r_{tt} of the

whole test. Where r_{XX} is the reliability coefficient of a test of unit length, k is the number of times this test is to be lengthened, and r_{kk} is the reliability coefficient of the lengthened test, the Spearman-Brown formula is as follows :

$$r_{kk} = \frac{kr_{XX}}{1 + (k-1)r_{XX}} ; \quad k = \frac{r_{kk}(1-r_{XX})}{r_{XX}(1-r_{kk})}$$

For the split-half method, $r_{XX} = r_{hh}$, $k = 2$, and $r_{kk} = r_{rr}$.

$$\therefore r_{rr} = \frac{2r_{hh}}{1+r_{hh}}$$

This extended r_{rr} of the whole test is called the *coefficient of internal consistency* because it measures the equivalence of contents and hence, the consistency of the two half-tests. It is higher than the test-retest r_{rr} and the equivalent-forms r_{rr} because a simultaneous administration of both half-tests under identical conditions lowers s_e^2 substantially. But such simultaneous administration of both half-tests may increase the "true variance" and the r_{rr} falsely by causing the half-test scores suffer from errors in the same direction. Moreover, arbitrariness in splitting may add an element of bias to the computed r_{rr} .

This split-half method is eminently suitable for power tests where all candidates get enough time to attempt every item (§ 10.3). In a power test with items arranged in an ascending order of difficulty, an *odd-even split* is often done by separating the scores of items, bearing odd and even serial numbers, into two series to represent separate half-tests, thus ensuring equivalent difficulties of the two halves.

The internal consistency coefficient may be used in rating when different test items explore the same attribute.

However, this method is not suitable for high-speed tests where all examinees cannot attempt all the items. But it can be used for

speed tests if the splitting of test items into two halves is done in terms of both time and item difficulty ; the two halves can then be administered to the same group, one immediately after the other, as two independently timed tests.

The split-half method is not suitable for heterogeneous tests because the items of such a test cannot be grouped into two equivalent halves. But it may be used for homogeneous power tests in case of shortage of time and nonavailability of an alternate form of the test.

4. Rational equivalence method :

This is another method used to measure the internal (item-item) consistency of a homogeneous test whose items are all estimating a single specific attribute. The r_{rr} computed by this method is also a *coefficient of internal consistency* and measures both homogeneity of test items and equivalence of contents.

In the rational equivalence method, the entire test is administered at a time to a group of individuals. As splitting of the test is avoided, no bias is introduced in the computed r_{rr} owing to arbitrary grouping of items in separate halves. Assumptions for this method include similar item difficulties, dichotomy (e.g., yes/no, 1/0) of test items, uniform and high item-total correlations of scores, and the dependence of total variance (s_t^2) on item variances and covariances. Such assumptions prevent its application to heterogeneous and speed tests.

Where n is the number of test items, s_t^2 is the variance of total test scores, p is the proportion of subjects answering an item correctly, q is the proportion of subjects answering it wrongly, and Σpq is the sum of item variances given by the product of p and q for each test item, r_{rr} is computed by the *Kuder-Richardson formula 20* (K-R 20) :

$$r_n = \frac{n}{n-1} \left(1 - \frac{\sum pq}{s_t^2} \right).$$

K-R 20 is actually based on *analysis of variance* (§ 11.2). It is a measure of the average item intercorrelation of the test and an estimate of the average of split-half correlations obtained from all possible splittings of the test.

r_n ranges from 0 to 1.00 in value. Lower and less accurate r_n is given by *Kuder-Richardson formula 21*, using the mean test score (\bar{X}):

$$r_n = \frac{n}{n-1} \left(1 - \frac{\bar{X}(n-\bar{X})}{ns_t^2} \right).$$

Example 10.5.2.

The reliability coefficient for a test of 20 items was found to be 0.48. What will be the reliability coefficient if the test is lengthened by 10 more items?

Solution :

$$r_{XX} = 0.48. \quad n = 20. \quad k = \frac{\text{final test length}}{\text{initial test length}} = \frac{n+10}{n} = \frac{20+10}{20} = 1.5.$$

$$r_{kk} = \frac{kr_{XX}}{1+(k-1)r_{XX}} = \frac{1.5 \times 0.48}{1+(1.5-1)0.48} = 0.58.$$

Example 10.5.3.

Find the coefficient of internal consistency for a whole test whose split half-tests are correlated by 0.52.

Solution :

$$r_{hh} = 0.52. \quad \therefore r_n = \frac{2r_{hh}}{1+r_{hh}} = \frac{2 \times 0.52}{1+0.52} = 0.684.$$

Example 10.5.4.

Use Kuder-Richardson formula 21 to compute the coefficient of internal consistency for a test consisting of 56 items where the mean and SD of all the scores amount to 32 and 10.5 respectively.

Solution :

$$n = 56; \quad \bar{X} = 32; \quad s_t = 10.5; \quad \therefore s_t^2 = (10.5)^2 = 110.25.$$

$$r_n = \frac{n}{n-1} \left[1 - \frac{\bar{X}(n-\bar{X})}{ns_t^2} \right] = \frac{56}{56-1} \left[1 - \frac{32(56-32)}{56 \times 110.25} \right] = 0.892.$$

Example 10.5.5.

In a test consisting of 20 items, the sum of the item variances (Σpq) and the SD of all the test scores amounted respectively to 8.3725 and 6.5. Calculate the coefficient of internal consistency of the test by Kuder-Richardson formula 20.

Solution :

$$s_t = 6.5 ; \quad \therefore s_t^2 = (6.5)^2 = 42.25. \quad \Sigma pq = 8.3725. \quad n = 20.$$

$$\therefore r_{tt} = \frac{n}{n-1} \left[1 - \frac{\Sigma pq}{s_t^2} \right] = \frac{20}{20-1} \left[1 - \frac{8.3725}{42.25} \right] = 0.844.$$

Example 10.5.6.

If a test consisting of 20 items has a reliability coefficient of 0.45, how long it should be made to have a reliability coefficient of 0.80 ?

Solution :

$$r_{XX} = 0.45 ; \quad n = 20 ; \quad r_{kk} = 0.80.$$

$$\therefore k = \frac{r_{kk}(1-r_{XX})}{r_{XX}(1-r_{kk})} = \frac{0.80(1-0.45)}{0.45(1-0.80)} = 4.9.$$

So the test should be made 4.9 times longer. Thus, the lengthened test should have the following number of test items :

$$nk = 20 \times 4.9 = 98 \text{ items.}$$

10.6 VALIDITY OF A TEST

Validity is the capacity of a test to measure and predict the specific variable under investigation, in exclusion of other variables. It indicates (i) the *relevance* of the test to the variable or trait to be investigated, (ii) the *discriminatory power* of the test to exclude the measurement of other variables, and (iii) the *predictive value* of the test for only the specific trait. A test for clerical aptitude is valid if it can predict success in clerical jobs, but gives no measure of any other variable like the aptitude for salesmanship. Validity is affected by *systematic errors of measurement* (page 251).

Validity should not be generalized beyond

the specific purpose or standard of a test. It needs empirical investigation. So, for constructing a valid test, (i) the attribute to be investigated should be precisely fixed, (ii) the test items should be carefully chosen to reveal and measure that attribute in exclusion of others, and finally (iii) its validity should be determined by correlating the test scores with some *external criterion*. The latter may be either an objective measure of the same attribute or any other variable representing the latter. External criteria differ according to the purpose of the test, because different tests require different types of validity. External criteria for intelligence tests include school marks, teacher's ratings, grades in public examinations, achievement test scores, and

similar but more elaborate tests of established validity like Binet-Simon and Wechsler scales. Achievement tests are validated against the criteria of actual courses of study, analysed and chosen by experts. The criteria for personality tests are the actual forms of behaviour, case histories, clinician's reports and comparison of test scores before and after therapy.

Validity is a measure of accuracy with which the test scores predict the variable to be explored by the test. It is ordinarily taken to be directly proportional to the magnitude of the *validity coefficient* (r_{XY} or r_{YX}) which is the correlation coefficient between the scores (X) of the given test and those (Y) of an external criterion. Validity varies with the nature of the group tested; a group with a wider ability range yields a higher r_{XY} than a small selected group.

Sometimes, different meanings of validity are expressed by terms like intrinsic and relevant validities.

(a) *Intrinsic validity* indicates the capacity of test scores (X_t) to denote true scores (X_∞). It is given by the *intrinsic validity coefficient* ($r_{t\infty}$) which is the square root of the reliability coefficient (r_{tt}): $r_{t\infty} = \sqrt{r_{tt}}$.

(b) *Relevant validity* shows the extent to which a test measures the factors common with another test. The *coefficient of relevant validity*

is the correlation coefficient of test scores with other measures, considering the common factors and eliminating the specific variance of the test scores. It is given by the square root of common factor variance in the scores.

Validity of a homogeneous test may be enhanced by increasing the test length without disturbing the homogeneity, because such an elongation either increases the proportion of the true variance s_∞^2 or adds new factors to increase the common factor loading of the test. The validity coefficient $r_{Y(kX)}$ of the elongated test depends upon the reliability coefficient (r_{XX}) of the test of unit length, the validity coefficient (r_{YX} or r_{XY}) of the same, and the number (k) of times the test has been elongated.

$$r_{Y(kX)} = \frac{r_{YX}}{\sqrt{\frac{1-r_{XX}}{k} + r_{XX}}}; \quad k = \frac{1-r_{XX}}{\frac{r_{YX}^2}{r_{Y(kX)}^2} - r_{XX}}$$

But unlike reliability, validity rises at a lower rate and never attains the perfect stage with the rise in test length. The *SE of estimate* (s_e) for r_{YX} measures the average random error in predicting individual criterion scores from test scores, owing to an imperfect validity of the test.

$$s_e = s_Y \sqrt{1 - r_{YX}^2}$$

Example 10.6.1.

An achievement test has a validity coefficient of 0.64 between the test scores and the grade performance, and a reliability coefficient of 0.55. What will be the expected validity if the test is lengthened to twice its original length? How many times it should be lengthened to get a validity of 0.80?

Solution:

$$r_{YX} = 0.64; \quad r_{XX} = 0.55.$$

(a) When the test is elongated to twice its original length,

$$k = 2, \quad \therefore r_{Y(kX)} = \frac{r_{YX}}{\sqrt{\frac{1-r_{XX}}{k} + r_{XX}}} = \frac{0.64}{\sqrt{\frac{1-0.55}{2} + 0.55}} = 0.73.$$

(b) To get a validity coefficient $r_{Y(kX)}$ of 0.80,

$$k = \frac{1-r_{XX}}{\frac{r_{YX}^2}{r_{Y(kX)}^2} - r_{XX}} = \frac{1-0.55}{\frac{(0.64)^2}{(0.80)^2} - 0.55} = 5.$$

So, the test should be lengthened to 5 times its original length for getting a validity of 0.80.

Types of validity

Three types of validity, viz., content validity, construct validity and criterion-related validity, have been recognized by the American Psychological Association. In addition to these, a fourth type of validity, called the job-component validity, has also been recognized in industrial psychology. The type of validity to be estimated is determined by the purpose of the test.

1. Content validity :

It is based on the logical analysis and proper sampling of the contents of the test. It is estimated to ensure the relevance of both the individual test items and the total test contents to the behavioral domain under consideration. It also investigates whether different aspects of the relevant behaviour are assessed in correct proportions by the test items, and how far those items and the cognitive processes involved form a representative sample of the variable to be measured.

Content validity thus assesses the suitability of a test in evaluating the existing status of an individual in a specific area of behaviour. It is normally used for educational achievement tests by computing the correlation coefficient

between the test scores (predictor) and some independent criterion for the pertinent variable. In work psychology, content validity is estimated while validating the tests for job performance. For this, the test items are examined, before actual application of the test, to find how far their contents are appropriate for the selection of personnel for the job. But content validity is not applicable to personality tests.

Content validity is worked out in the following ways.

(i) When both the test scores (X) and the criterion scores (Y) constitute continuous variables, *content validity coefficient* is given by the product-moment r_{XY} between the two.

(ii) When correlating continuous test scores with a dichotomized criterion, either point biserial r_{pbi} or biserial r_b is computed according to the genuine or arbitrary nature of the dichotomy of the criterion (pages 173-180).

(iii) When both the criterion and the test scores are dichotomized, either phi coefficient or tetrachoric r is computed between them according to the genuine or arbitrary nature of the dichotomy (pages 181-185).

(iv) In case of more than one test as the

predictors, validity is given by the *multiple correlation coefficient* ($R_{1,23,\dots,m}$) between the combined test scores and the criterion (pages 158-161).

A test should not be taken as valid merely because its contents seem apparently to measure the ability being investigated. Such an appearance, often termed *face validity*, is considered as only one aspect of content validity ; it cannot be subjected to any statistical treatment and may merely serve in motivating the candidate as well as the investigator, and is sought for this purpose in professional selection tests and achievement tests of adults.

2. Construct validity :

Psychological construct is a hypothetical concept of such an abstract quality or attribute like intelligence, emotional stability, mechanical ability and inner drive, as cannot be observed directly, but can be measured from the behavioral expressions.

Construct validity is estimated to assess the suitability of a test in evaluating the status of a person in a given abstract area of behaviour called a psychological construct. It may also be worked out for attitude scales and personality tests.

Construct validity may also be estimated for the validation of ratings such as those of job characteristics ; however, an appropriate external criterion is often hard to get for the construct validation of ratings.

For construct validation, each test item should be chosen with a clear understanding of the psychological construct so as to reflect, detect and measure the latter. The performance in the test is subsequently judged by its agreement with the psychological construct. Construct validity is ensured if the scores in the given test have either a significant

moderately high positive correlation with the scores in other tests known to measure the same trait satisfactorily (*convergent validity*), or a negligible or negative correlation with the scores in tests for dissimilar traits (*discriminant validity*). For example, the construct validity of an intelligence test is ensured either by the significantly high positive correlation of its scores with those of already proven intelligence tests or sometimes, by their poor correlation with those of musical or other special ability tests. Correlation is computed here between the scores of the test under investigation and *more than one criterion*.

The *construct validity coefficient* depends on the sharing of some common factors by the test and the criteria ; hence, *factor analysis* (§ 10.8) may be undertaken for construct validity.

3. Criterion-related validity :

This includes predictive and concurrent validities. For both, scores of the test (predictor) are usually correlated with scores of an independent criterion by computing either Spearman's rho or Person's product-moment r .

In validating tests for personnel selection, a measure of job performance is usually chosen as the criterion.

(a) *Predictive validity* : It is given by the correlation between the scores of the test for predicting a particular trait, and those of a criterion that is a measure of a specific *subsequent performance*. Its main objective is a prediction of future performance. Selection or classification of candidates can be done on the basis of test scores supposedly reflecting the future performance. For example, the scholastic aptitude test for admission to an academic institution is validated against subsequent academic achievements in the school. The predictive validity of a test for personnel selection is worked out by correlating the test scores, collected at the time of appointment,

with the scores of a job performance test administered subsequently. The accuracy of prediction is indicated by the *coefficient of predictive validity* which is the correlation coefficient between the test scores and the criterion scores. *Expectancy tables* are frequently used for predictive validity (§ 10.7). The computation of predictive validity involves a follow-up study because validation is done here against a criterion of subsequent performance. For perfect predictive validity, the criterion should be highly reliable and self-correlated, enabling the generalization from the scores of one variable to those of another. For many selection tests, a moderate correlation serves the purpose. A test for predictive validity should have good content and construct validities for its perfect application.

(b) *Concurrent validity* : It is given by the correlation of a test with an *existing* criterion of, say, job performance instead of a subsequent performance. So, it needs less time and labour. For example, the concurrent validity of an arithmetic test is given by the correlation coefficient between the test scores and the already achieved class grades of the tested individuals. Similarly, the scores of any intelligence test can be validated against the scores already obtained in the Stanford-Binet scale. Thus, concurrent validity shows the validity of the test as a measure of the *present status* of an individual while predictive validity assesses the test as a measure of his *future status*.

When both the predictor and the criterion are perfectly reliable, criterion-related validities tend to reach maximum magnitudes, approaching but rarely attaining the square root of the reliability coefficient of the test.

4. Job-component validity :

The concept of this validity is based on the assumption that the requirements for a given job should have some components common or

comparable with those of a test for performance in that job. When the relevant component is a specific job skill, it refers to *content validity* of the test and is estimated as such. In case of a basic attribute, on the contrary, it may be estimated as *construct validity*.

Relation between reliability and validity

(a) A test is far from valid if it cannot measure a trait reliably. A reliable test should also be theoretically valid because proper item selection and elongation of tests lead to increased proportions of true variances and hence, to enhanced common factor loadings. However, reliability does not always ensure validity in practice.

(b) Reliability depends on the proportion of *true variance* in a test while validity depends basically on the *common factor variances* which, however, form a component of the true variance.

(c) Reliability is a measure of self-correlation of a test and depends on the *identical difficulty levels* and *high internal consistency* of test items. But predictive validity depends on *differing difficulty levels* of test items and their *low correlations*. So, both reliability and validity may not be uniformly high for a test.

(d) Lengthening a test increases the proportion of common factor variances and hence, the proportion of true variances also. Rise in the number of common factors, shared by the test and the criterion, increases the validity while the rise in the proportion of true variance enhances reliability. But unlike reliability, validity cannot be raised to the maximum level by the increase in the test length alone.

(e) Reliability may be low in a heterogeneous test with its items measuring different factors, but a high predictive validity may still

result from the common factors shared by the test items and the criterion. Reliability may be high in a homogeneous test with high internal consistency. But validity may often still be low due to the dearth of common factors for the items and the criterion.

Tests should possess sufficient validities along with reasonable degrees of reliability. A test serves no purpose if it gives consistent results on repetition, but fails to measure the desired trait. Both reliability and validity may be reasonably ensured by replacing a single test with a test battery of heterogeneous type, but consisting of individual tests of homogeneous natures.

10.7 EXPECTANCY TABLE

An expectancy table is constructed from a scatter diagram or correlation table giving the bivariate distribution of two continuous variables (pages 150-152). It expresses the relationship between the test scores (predictor) and the scores of the criterion. It indicates the probability of attaining a particular grade or class of performance when the test scores are known. Where both the test scores and the criterion scores constitute continuous variables, the paired scores of the two for the tested individuals are arranged in a two-way *bivariate*

frequency distribution which is used as the *correlation table* (pages 150-152). Each cell frequency of such a correlation table or scatter diagram (Table 10.5) is converted to a percentage of the marginal total f_r of the row to which that cell belongs (Table 10.6). Expectancy tables may be constructed even with more than one predictor, when necessary, or with two classes of a dichotomized criterion.

10.8 FACTOR ANALYSIS

Factor analysis is the right statistical procedure for *construct validity* of tests. If scores of a test correlate highly with scores of another similar and already validated test or criterion measuring the same attribute, the two tests may have common factors shared by their scores (*convergent validity*). The poor correlation between the scores of two tests, measuring dissimilar attributes, may result from their poor loading with common factors (*discriminant validity*). Thus, factor analysis can explain the intercorrelations between tests by analyzing the common factors.

Factor analysis identifies the factors or psychological components in a test, assesses their relative independence, correlates them individually to a criterion for determining their individual contributions to the total test scores,

Table 10.5. Correlation table for algebra test scores and class examination marks in mathematics.

Algebra test scores	Maths marks						f_r
	21-30	31-40	41-50	51-60	61-70	71-80	
66-70				2	2	5	9
61-65			2	4	5	1	12
56-60		3	3	4	5	1	16
51-55		5	4	5	4		18
46-50	1	4	4	5	2		16
41-45	2	4	2				8
36-40	3	2	1				6
31-35	5						5
Grades	F	E	D	C	B	A	90

and analyzes the weights of factors responsible for the common variance of scores in two tests. It thus searches for common factors, shared by two tests and responsible for their correlation. Common factors can explain the inter-correlation between a group of tests. For example, tests for arithmetic problems, number completion, addition and multiplication may show high positive correlations with each other due to a common 'numerical factor' in all of them; but these tests show low and insignificant correlations with vocabulary and sentence completion tests, loaded with a 'verbal factor' instead of the 'numerical' one. Factor validity is given by the loading of a test with factors and its correlation with each factor. So, factor analysis validates a test in the form of factor loading or intercorrelation between each test and the factor. It is usually applied to the data of interdependent variables, not differentiated into dependent and independent ones. It identifies the most specific ability required for a particular task and reduces the number of original test items to a smaller number of common factors sufficient to explain the correlation between the tests.

A good criterion can be ensured by its factor analysis. This intercorrelates criterion measures among themselves as also correlates

them with validated tests of anticipated common factors. Investigations of factor loadings of the criterion help to choose the most relevant and representative criterion, to include the most predictive tests in a test battery, to give weights to criteria for combinations, and to detect the insufficiency of a proposed test battery in predicting any factor of the criterion.

Of the numerous theorems of factor theory, theorems I and II suffice to explain validity.

Theorem I of factor theory :

True variance (s_{∞}^2) of a test score is the sum of (i) common factor variances ($s_a^2, s_b^2, \dots, s_k^2$) for the common factors (a, b, \dots, k) shared by many tests, and (ii) a specific factor variance (s_i^2) of the given test, which is shared only by its parallel and equivalent forms.

$$s_{\infty}^2 = (s_a^2 + s_b^2 + \dots + s_k^2) + s_i^2.$$

The total variance s_i^2 equals the sum of s_{∞}^2 and the error variance s_e^2 . Where a_x^2, b_x^2 , etc., are the proportions of s_i^2 due to the common factors, s_x^2 or specificity is the proportion of s_i^2 in s_i^2 , and e_x^2 is the proportion of s_e^2 in s_i^2 ,

$$\begin{aligned} s_i^2 &= s_{\infty}^2 + s_e^2 \\ &= (s_a^2 + s_b^2 + \dots + s_k^2 + s_i^2) + s_e^2 ; \end{aligned}$$

Table 10.6. Expectancy table for grades in mathematics on algebra test scores of Table 10.5.

Algebra test scores	Percentages of cases in grades				
	F	E	D	C	B
66-70				22	22
61-65			17	33	42
56-60		19	19	25	31
51-55		28	22	28	22
46-50	6	25	25	31	13
41-45	25	50	25		
36-40	50	33	17		
31-35	100				

$$\begin{aligned}\text{or, } 1.00 &= \left[\frac{s_a^2}{s_t^2} + \frac{s_b^2}{s_t^2} + \dots + \frac{s_k^2}{s_t^2} + \frac{s_s^2}{s_t^2} \right] + \frac{s_e^2}{s_t^2} \\ &= (a_X^2 + b_X^2 + \dots + k_X^2 + s_X^2) + e_X^2 ; \\ \text{or, } 1 - e_X^2 &= (a_X^2 + b_X^2 + \dots + k_X^2) + s_X^2 ; \\ \text{or, } \frac{s_{\infty}^2}{s_t^2} &= (a_X^2 + b_X^2 + \dots + k_X^2) + s_X^2.\end{aligned}$$

Where h_X^2 or *communality* is the sum of proportions of common factor variances in the test scores, and u_X^2 or *uniqueness* is the sum of proportions of specific and error variances in the test scores,

$$\begin{aligned}r_{tt} &= \frac{s_{\infty}^2}{s_t^2} = a_X^2 + b_X^2 + \dots + k_X^2 + s_X^2 ; \\ \text{or, } r_{tt} - s_X^2 &= h_X^2 = a_X^2 + b_X^2 + \dots + k_X^2 ; \\ \therefore h_X^2 &= \frac{s_{\infty}^2}{s_t^2} - s_X^2 = \frac{s_{\infty}^2}{s_t^2} - \frac{s_e^2}{s_t^2}, \text{ and} \\ u_X^2 &= 1 - h_X^2 = s_X^2 + e_X^2.\end{aligned}$$

Factor loadings are the square roots of the proportions of common factor variances ; each factor loading, viz., a_X , b_X , etc., is the correlation coefficient between the relevant common factor and the total test score. This correlation coefficient, called the *factor validity*, is an estimate of the capacity of the test to measure the trait underlying the given factor.

Theorem II of factor theory :

Validity coefficient is the correlation coefficient (r_{XY} or r_{YX}) between the test (X) and either an external criterion or another validated test (Y). It equals the sum of crossproducts of common factor loadings of X and Y, because r_{XY} results from common factors shared by them and amounts to 0 in absence of such common factors. Construct validity coefficient can thus be computed from factorial validities or common factor loadings.

$$r_{XY} \text{ or } r_{YX} = a_X a_Y + b_X b_Y + \dots + k_X k_Y.$$

10.9 STANDARDIZATION AND NORMS

The final stage in test construction consists of the standardization of the test. A test has been standardized if (i) its items have been properly analyzed and chosen, (ii) procedures of administration and scoring have been made uniform, (iii) instructions for its application as also the scoring keys have been provided, and (iv) norms have been established and tabulated for interpreting the test scores.

A raw score can be made meaningful and significant by comparing it with a standard. For example, numerical scores of 125, 137 and 144 in the US Army Alpha Group Intelligence Test indicate only the relative positions of the tested individuals, and acquire real significance only when judged against the score distribution of a representative sample. To achieve this, the establishment of a norm is essential for each sample.

A *norm* is the average score of a representative group in terms of a convenient scale of converted or transformed scores. Raw scores are, therefore, frequently converted into transformed scores by either linear or nonlinear transformation. Transformed scores help the investigator to compare an individual's performance with those of other persons and with his own performance in different tests by rendering his relative positions clearly comparable with the standardization sample. Usually expressed in the same unit for different tests, transformed scores as well as their means, dispersions and forms of distribution are comparable. The essential criterion for establishing norms is the representative nature of the sample rather than its size. Norms are limited to the particular group or population from which the representative sample was drawn when establishing the norm. Local norms, group norms, regional norms or national norms are established separately according to the nature of the sample. Even the established norms may get outdated and need updating.

Transformations of raw scores

Raw scores belong to an arbitrary scale. So, they need transformation into a suitable common scale for gaining a standard meaning, a common reference value and a comparability with other test scores. Raw scores need transformation also for ensuring additivity of the treatment effect, and for making the data normally distributed and homoscedastic (homogeneous with respect to variances) — this may fulfil the assumptions for statistical tests like the *t* test and anova. Raw scores of psychological tests are frequently converted to (a) percentile scores, (b) age scores, (c) ratio IQ and (d) standard scores, and their variations like *z* scores, *T* scores, *C* scores, stanines and deviation IQ.

Linear transformations :

Linear transformations of raw scores merely change the zero point of the scale and/or the unit of measurement. So, they change the mean and *SD* of the raw scores without altering the original shape and properties like skewness and kurtosis of their distribution. The differences between raw scores correspond closely in relative magnitude to those between the respective linearly transformed scores (§ 5.4). The *z* and *t* scores are examples of linearly transformed scores.

Nonlinear transformations :

Nonlinear transformations change not only the mean and the *SD*, but also the shape, skewness and kurtosis of the original raw score distribution. They are often tried for converting a non-normal distribution of raw scores into a normal distribution of transformed scores. Percentile scores, *T* scores, *C* scores, stanines and mental age scores are examples of nonlinear transformations of raw scores in psychology.

Percentile or centile scale

Conversion of raw scores into percentile

ranks or *PR* (page 42) is a form of nonlinear transformation frequently used during the standardization of psychological test scores. The percentile scale is a *transformed ordinal scale*. It gives a rectangular form to the frequency distribution of the relevant variable without affecting the rank orders of original scores. It expresses individual scores in terms of a typical sample of 100. Each interval of this scale contains an equal number of cases. A *percentile point* is a value, below which lies the stipulated percentage of cases (page 40). If 70% of the individuals of a sample have scored below 40 in a test, then 40 is the 70th percentile point while 70 is the *PR* of the score 40 (page 42). Percentile points of the original scale are transformed into the corresponding *PR* values to make the scores of an individual in different tests comparable to each other. Thus, a relative measure of his status in different traits can be obtained from his *PR* values in different tests.

Advantages : The percentile scale is (a) easy to compute, comprehend and interpret, (b) applicable to any type of ability or personality test without complicated statistics, and (c) applicable to both normal and non-normal distributions.

Limitations : This scale (a) shows only the relative rank of an individual in a representative sample and cannot indicate the actual difference between the test scores of two individuals, (b) is unsuitable for small samples, (c) is not useful for many subsequent statistical computations, and (d) shows identical differences in σ units as unequal differences (Fig. 10.2). For normally distributed raw scores, the difference between the *PR*s of 50 and 16 amounts to 34, but that between the *PR*s of 16 and 2 amounts to 14 only, although both differences represent a difference of 1σ . This is because larger differences near the tail ends of the normal curve are reduced and smaller differences near its centre are exaggerated in the percentile

scale. These limitations often make the z scores preferable to *PR*.

Age scores

The commonly known age score is the mental age score of Binet-Simon Age Scale of Intelligence (1908) and its revisions. Here, the test items are specified for each age level and arranged in an ascending order of difficulty. The test measures the intellectual level of a child in terms of the average chronological age of normal children with the same intellectual standard. The intelligence level of the child is expressed as the *mental age*. The latter is defined as the age level of normal children whose average performance in the test coincides with that of the tested child. If a child, aged 8, answers successfully all the test items placed at the age level 8, (i.e., the items correctly answered by a large number of normal children of 8 years during the grading of items), then he is assigned an *initial level* of 8 years. The child is then assigned an *additional credit* of 1 year each, for passing all the items at each successive higher age level; but *no partial credit* is given to him for passing less than all the items at any age level. The sum of the initial level and all the additional credits gives his *mental age* in the 1908 scale and its 1911 revision. The average child of each age group has the same mental age as his chronological age. This original procedure was, however, considered unsound because of the absence of partial credits for part successes at any age level.

In the 1916 Stanford Revision of Binet Scale by L. M. Terman and Maud Merrill, provisions were made for the award of partial credits. The *highest age level*, where the child passed all the test items, is taken as the starting point or *basal age* for that child. Then, a stipulated number of months is credited to him for each item passed above that level. The age level, where the child can answer none of the

items, is called the *ceiling* or *apical age*. The sum of the basal age and the additional fractional credits gives his *mental age*.

Age norms are differently used in many other tests, not arranged this way according to age levels. Scores are awarded there either on the total number of items passed or on the total time taken. The mental age is then fixed with reference to the mean score of average children of corresponding age level in the normative sample.

Ratio IQ

Mental age score itself carries a significance. But (i) it shows only the level of intellectual maturity of a child and cannot differentiate between average, bright and dull children because of no reference to their chronological age. (ii) The unit does not remain stable with advancing age — the difference in mental age between earlier years of life is greater than that between the later years, because the rate of intellectual development progressively declines with age. To remedy these limitations, Terman and Merrill converted the mental age score into a ratio called the *intelligence quotient* (IQ) or *ratio IQ* in their Stanford revision (1916) of the Binet scale. Where MA and CA represent respectively the mental age and the chronological age of a child,

$$IQ = \frac{MA}{CA} \times 100.$$

IQ is a *quotient norm* or *ratio norm* which permits the interpretation of mental age scores relative to the chronological age. It thus places ability in a correct perspective with regard to chronological age. Evidently, 100 will be the average IQ for each age group. IQs above and below 100 indicate respectively the acceleration and the retardation of intellectual performances.

IQ is a *nonlinearly transformed age score*

— the transformation approximately normalizes the score distribution with a mean of 100 and *SD* of 16; it thus changes the original rank orders of mental age scores as well as the mean, *SD* and shape of the original mental age distribution.

Limitations : (a) Age scores including IQ cannot assess other traits which do not change with the chronological age. (b) Use of the chronological age in its computation poses a problem in determining IQs of older people with supposedly sloped intellectual developments — IQ goes on declining with age as mental age attains its peak at maturity. To avoid this, 16 was fixed as the highest age with complete intellectual growth. But IQ may still fluctuate. (c) IQ as defined in the 1937 revision of Stanford-Binet scale shows agewise and testwise variations of its *SD*. This may be remedied by using either a correlation table for each age level or, still better, the deviation IQ discussed later.

Standard scores

Standard scores are either *linearly* or *nonlinearly* transformed raw scores. The simplest form of linearly derived standard score is the *z* score (page 74). Being a form of linear transformation, *z* scores bear the same numerical relations and relative differences between each other as those between the raw scores from which they have been derived. Transformation into *z* scores cannot, therefore, normalize a non-normal raw score distribution and only alters the mean and *SD* of the latter.

There are two limitations of *z*. (i) A raw score below the mean gives a negative *z* score. (ii) The unit of 1σ is relatively large and often introduces decimals in the computed *z*. To avoid these limitations, *transformed z scores* (*z'*) or *proper standard scores* are computed by multiplying *z* scores with a constant number much higher than 1, and either adding the

product to or subtracting it from another constant much higher than 0. For such further transformations, the mean is usually taken as 50 or 500, and the *SD* as 3, 10, 15 or 50. Such transformed standard scores include the *T* scores ($\bar{X}_s = 50, s_s = 10$), the Scholastic Aptitude Test (SAT) scores ($\bar{X}_s = 500, s_s = 100$), the Graduate Record Examinations (GRE) scores ($\bar{X}_s = 500, s_s = 100$), Wechsler Block Design sub-test norms ($\bar{X}_s = 10, s_s = 3$), Differential Aptitude Test (DAT) scores ($\bar{X}_s = 50, s_s = 10$), and Wechsler deviation IQ scores ($\bar{X}_s = 100, s_s = 15$). For the SAT scores, for example,

$$z' = \bar{X}_s + s_s z = 500 + 100z.$$

Standard scores are converted to *normalized standard scores* by transforming raw scores into equivalent points on a normal distribution. Each such point corresponds to the same level of the variable as the respective raw score. For example, raw scores are converted to percentile ranks, and the latter to the corresponding *z* scores of a normal distribution. Normalized standard scores have the same mean of 0 and *SD* of 1 as those of the linearly derived standard scores. The normalized standard score can be transformed into *T* score, *C* score, stanine, deviation IQ, etc., by multiplying it with a convenient number and adding/subtracting that product to/from another desired number.

T scale

T scores are *normalized standard scores* computed by the nonlinear transformation of raw scores. The transformation of raw scores to *T* scores changes even a non-normal raw score distribution into a normal one, but leaves the rank orders of the scores unaltered (Fig. 10.2). The centre (mean) and *SD* of the *T* score distribution amount to 50 and 10 respectively. The *T* scale ranges from 0 (-5σ) to 100 ($+5\sigma$)

and the unit of measurement amounts to 0.1σ .

Computation of *T* scores :

(a) The raw scores are grouped into class intervals whose midpoints (X_c) are also computed (page 16). The frequencies (f) of raw scores are entered in the respective class intervals (Table 10.7).

(b) For each class interval, the cumulative frequency (cf) upto its lower limit (X_l) is computed by adding the frequencies of all class intervals below it ; for example, for the interval 36-40 of Table 10.7, $cf = 3 + 8 = 11$.

(c) For each interval, the cf upto its X_c is computed by adding half the frequency of that interval to the cf upto its X_l . For example, the cf upto the X_c of 36-40 amounts to : $11 + \frac{1}{2} \times 12 = 17$ (Table 10.7).

(d) For each interval, the cumulative proportion (cp) upto its X_c is computed by dividing the cf upto its X_c with the sample size n . For example, for the class interval 36-40 of Table 10.7, $cp = 17.0 \div 100 = 0.17$. The cp represents the corresponding *PR* expressed as a proportion.

(e) 0.50 is deducted from each cp and the z score for that obtained "area" is recorded from

Table A of unit normal curve given in the Appendix. This z score would correspond to the upper limit of the fractional area represented by the corresponding cp . For example, 0.5000 is deducted from the cp of 0.3050 to give a sum of -0.1950 ; the z score corresponding to this fractional area of -0.1950 amounts to -0.51 (Table A). The z scores for all the cp values are found out in this way and entered against the respective midpoints.

(f) Each z score is then converted to the corresponding *T* score as follows :

$$T = 50 + 10z$$

For example, for the z score of -0.51 ,

$$T = 50 + 10z = 50 + 10 \times (-0.51) = 44.9$$

T scores are obtained in this way for all the midpoints of the class intervals.

Advantages of *T* scale : (a) *T* scale can change some non-normal raw score distributions into normal distributions. (b) It leaves the rank order of the raw scores unaltered. (c) Its mean, *SD* and unit are convenient for using it even beyond the given range of the population.

Limitations of *T* scale : (a) Its unit of 0.1σ

Table 10.7. Computation of *T* scores for a raw score distribution.

Class intervals	X_c	f	cf upto X_l	cf upto X_c	cp upto X_c	z	T
66-70	68	2	98	99.0	0.990	+ 2.33	73.3
61-65	63	5	93	95.5	0.955	+ 1.70	67.0
56-60	58	10	83	88.0	0.880	+ 1.18	61.8
51-55	53	20	63	73.0	0.730	+ 0.61	56.1
46-50	48	25	38	50.5	0.505	+ 0.01	50.1
41-45	43	15	23	30.5	0.305	- 0.51	44.9
36-40	38	12	11	17.0	0.170	- 0.95	40.5
31-35	33	8	3	7.0	0.070	- 1.48	35.2
26-30	28	3	0	1.5	0.015	- 2.17	28.3
Total		100 (n)					

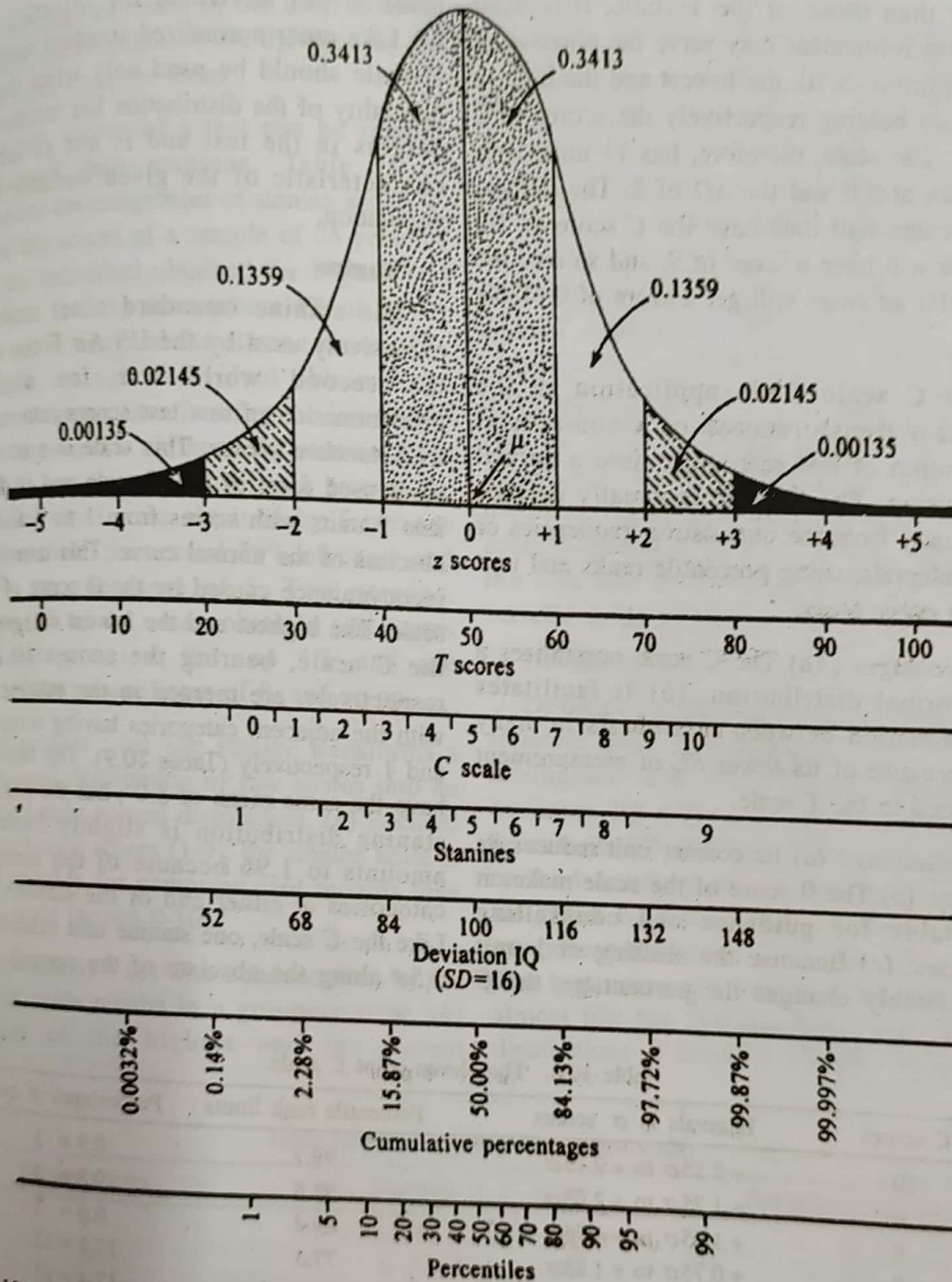


Fig 10.2. Relations between different types of standard scores, cumulative percentages and percentiles.

is too small to serve the purpose of many tests. from its unit.

(b) It has a high *SE* of measurement, making it difficult to distinguish between individuals by using 0.1σ as the unit. (c) The scale may not give as accurate measurements as appears

C scale

The C scale is another scale of *normalized standard scores*. Because its units (0.5σ) are

coarser than those of the T scale, it is used when less refinement may serve the purpose. It ranges from 0 to 10, the lowest and the highest categories bearing respectively the scores of 0 and 10. The scale, therefore, has 11 units with the mean at 5.0 and the SD of 2. The highest 1% of cases will thus have the C score of 10, next 3% will have a score of 9, and so on; the lowest 1% of cases will get a score of 0 (Table 10.8).

The C scale finds application in the nonlinear transformation of a non-normal distribution of raw test scores into a normal distribution. The C scale is usually derived graphically from the cumulative frequencies of class intervals, using percentile ranks and unit normal curve areas.

Advantages : (a) The C scale normalizes a non-normal distribution. (b) It facilitates discrimination between individuals in many tests because of its lower SE of measurement compared to the T scale.

Limitations : (a) Its coarser unit reduces its efficacy. (b) The 0 score of the scale makes it unsuitable for guidance and counselling purposes. (c) Because the shifting of 1 unit considerably changes the percentage, the C

scale is not advisable for selection tests. (d) Like other normalized standard scores, the C scale should be used only when the non-normality of the distribution has resulted from defects in the test and is not an inherent characteristic of the given variable in the population.

Stanine

The stanine (standard nine) scale was extensively used by the US Air Force during the second world war for nonlinear transformation of raw test scores into *normalized standard scores*. This scale is a somewhat condensed form of the C scale and is divided into 9 units with scores from 1 to 9 along the abscissa of the normal curve. This removes the inconvenience caused by the 0 score of the C scale. The highest and the lowest categories of the C scale, bearing the scores 10 and 0 respectively, are merged in the stanine scale with the adjacent categories having scores of 9 and 1 respectively (Table 10.9). The two scales have the same mean of 5.0; but the SD of the stanine distribution is slightly lower and amounts to 1.96 because of the merger of categories at either end of the stanine scale. Like the C scale, one stanine unit extends over 0.5σ along the abscissa of the normal curve.

Table 10.8. The eleven-point C scale.

C scores	Intervals in σ scores	Percentile rank limits	Percentages of cases
10	+ 2.25 σ to + 2.75 σ	99.7	0.9 \approx 1
9	+ 1.75 σ to + 2.25 σ	98.8	2.8 \approx 3
8	+ 1.25 σ to + 1.75 σ	89.4	6.6 \approx 7
7	+ 0.75 σ to + 1.25 σ	77.3	12.1 \approx 12
6	+ 0.25 σ to + 0.75 σ	59.9	17.4 \approx 17
5	- 0.25 σ to + 0.25 σ	40.1	19.8 \approx 20
4	- 0.75 σ to - 0.25 σ	22.7	17.4 \approx 17
3	- 1.25 σ to - 0.75 σ	10.6	12.1 \approx 12
2	- 1.75 σ to - 1.25 σ	4.0	6.6 \approx 7
1	- 2.25 σ to - 1.75 σ	1.2	2.8 \approx 3
0	from below - 2.75 σ to - 2.25 σ	0.3	0.9 \approx 1

The percentage of cases in the 0.5σ interval falls with the increase in the distance of the stanine score from the mean.

The raw scores of a test can be readily transformed into stanines. Table 10.10 summarizes the assignment of stanine scores to the raw test scores of a sample of 25 persons. Only one individual obtained the highest test score, thus forming the topmost 4% of the 25 cases, and is assigned the stanine score of 9. Similarly, a single individual obtained the lowest test score, thus forming the lowest 4% of the cases and is assigned the stanine score of 1. The test scores immediately around the mean raw score were obtained by 5 persons who constitute the central 20% of cases and get the stanine score of 5. Two persons, forming 7% of the cases, obtained test scores immediately below the top 4% and are assigned the stanine score of 8, and so on.

Advantages : (a) The stanine transforms a non-normal distribution of raw scores into an approximately normal distribution. (b) Because of single-digit scores (1-9), the stanine scale is more suitable for computer card records than the C scale (scores 0-10).

Limitations : (a) The condensation in the stanine scale results in a grouping error. (b) Merger of the highest and the lowest

categories, corresponding to the C scores of 10 and 0, with the respective neighbouring categories deprives the stanine scale of the power to discriminate between the highest 1% or the lowest 1% of cases and those immediately adjacent. This affects guidance work. (c) The grouping in the stanine scale lacks fineness and precision ; so, the transformed scores form only *approximately* normal distribution. Thus, z scores should suffice as the standard scores for a near-normal distribution of raw test scores and stanines need not be computed there. (d) Like other normalized standard scores, stanines should be used only when the non-normality of the raw score distribution has resulted from defects in the test and is not a characteristic of the variable in the population.

Deviation IQ

Deviation IQ is another form of *normalized standard score*. It was first used in Wechsler Intelligence Scale (1939). Unlike the ratio IQ, deviation IQ suffers no agewise fluctuation because its SD does not change with age. The method consists of the computation of a weighted IQ from each raw score and the assignment of a mean of 100 and an SD of 15. The test users can interpret the test scores almost like the Stanford-Binet ratio IQ. The distributions of deviation IQ (Fig. 10.2) and of

Table 10.9. The nine-point stanine scale.

Stanine scores	Intervals in σ scores	Percentages of cases
9	+ 1.75 σ to + 2.25 σ and above	4
8	+ 1.25 σ to + 1.75 σ	7
7	+ 0.75 σ to + 1.25 σ	12
6	+ 0.25 σ to + 0.75 σ	17
5	- 0.25 σ to + 0.25 σ	20
4	- 0.75 σ to - 0.25 σ	17
3	- 1.25 σ to - 0.75 σ	12
2	- 1.75 σ to - 1.25 σ	7
1	from below - 2.25 σ to - 1.75 σ	4

Table 10.10. Assigning stanine scores to raw test scores of a sample of 25 persons.

Stanine scores	Percentages of cases	Calculated frequencies	Equivalent raw scores
		1	61
9	4	2	58, 59
8	7	3	54, 55, 57
7	12	4	49, 51, 52, 53
6	17	5	43, 44, 44, 45, 48
5	20	4	38, 40, 41, 41
4	17	3	36, 37, 37
3	12	2	34, 35
2	7	1	33
1	4		
Total	100 %	25 (n)	

the 1937 revision of Stanford-Binet ratio IQ possess an identical mean and similar *SDs* (nearly 16). Deviation IQ has also been used in the L-M form of the 1960 revision of Stanford-Binet Scale, having a mean of 100 and an *SD* of 16. The transformation into deviation IQ gives the IQ a relative significance identical for all ages, and thereby makes the interpretation easier and less erroneous. Deviation IQ is

particularly applicable to intellectually mature persons older than 16 years.

The fourth revision of Stanford-Binet Scale by Thorndike et al in 1986 used the composite Standard Age Scores (SAS) having a mean of 100 and an *SD* of 16. They are also normalized standard scores expressed in the same unit as the deviation IQ of 1960.

GLOSSARY

age scores : mental age scores of Binet-Simon Age Scale of Intelligence or its subsequent revisions, which express the intelligence levels of children.

C scale : an eleven-point scale of normalized standard scores, obtained by nonlinear transformations of raw test scores and forming a distribution ranging from 0 to 10 with a mean of 5.0 and an *SD* of 2.

coefficient of equivalence : a measure of the equivalence of two parallel tests, that is obtained by correlating their scores on administering them to the same group of subjects either simultaneously or immediately after one another, and serves as an estimate of reliability by the alternate-forms method.

coefficient of internal consistency : a measure of the equivalence of contents and consistency of two half-tests into which a test has been split for estimating its reliability by the split-half method.

coefficient of stability and equivalence : a measure of both equivalence of contents of two tests and dependability of measurement over a time interval, that is obtained by correlating the scores of two equivalent or parallel forms of a test administered to the same group of subjects at two different times and serves as an estimate of reliability by the alternate-forms method.

concurrent validity : suitability of a test in evaluating the present status of a subject with respect to an existing criterion, as given by the correlation of the test score with his/her already achieved grade or score in the criterion.

construct validity : suitability of a test in evaluating the status of a subject in a given abstract behavioral area.

content validity : relevance of both the individual test items and the total test contents of a test to the given behavioral domain, and suitability of the test in evaluating the status of a subject in that behavioral area.

convergent validity : suitability of a test in measuring a trait, as given by a significant positive correlation of the test scores with the scores in other tests known to measure the same trait satisfactorily.

criterion : either a dependent variable whose likely scores are sought to be predicted from the scores of a test, or a dependent variable whose anticipated changes on exposure to a given independent variable are explored in an experiment.

dependent variable : the behavioral response which is either studied in an experiment on exposure of a group of subjects to different levels of an independent variable, or predicted from the scores of an independent variable in a test.

deviation IQ : normalized standard scores which are weighted IQ scores, and form a distribution having a mean of 100 and an *SD* of either 15 (Wechsler scale) or 16 (L-M form of Stanford-Binet scale).

difficulty value : the level of difficulty in answering a test item, estimated by the proportion of right answers to it by the subjects of a sample in the relevant test.

discriminant validity : suitability of a given test in measuring a trait, as shown by negligible or negative correlations of its scores with the scores of the same group of subjects in tests for dissimilar traits.

discriminatory value : capacity of a test item to discriminate between the subjects with respect to their answers to that item.

error of measurement, random : errors from such uncontrolled factors as erroneous scoring and interpretation, and affecting the reliability of a test.

error of measurement, standard : a measure of the random errors of measurement, estimated as the square root of the error variance of the test scores.

error of measurement, systematic : either overestimation or underestimation of the test scores of all the subjects systematically due to factors like defective test construction, and affecting the validity of the test.

expectancy table : table constructed from scatter diagram or correlation table for expressing the relation between the scores of a given test and the scores of a criterion.

extraneous variable : any variable which, though neither the dependent nor the independent variable in an experiment, still occurs in the experimental situation, in the experimental procedure or in the subjects used in the experiment.

factor analysis : statistical analysis for identifying and assessing the factors determining the psychological or physical characteristics of subjects in a test, for correlating the factors individually to a criterion, and for exploring the construct validity of the test.

independent variable : the variable deliberately chosen and used by the investigator either to study its effect on the dependent variable, or to predict the likely value of the criterion in a test.

intervening variable : any unobserved and hypothetical variable, presumably interposed between independent and dependent variables of an experiment, and affecting the dependent variable.

irrelevant variable : any variable that happens to occur in the subjects, situation or procedure of an experiment or test, but is neither used as the independent variable nor affects the dependent variable or criterion.

- item analysis** : statistical analysis of the test items of a psychological test to determine their difficulty levels and discriminatory powers.
- Kuder-Richardson formula 20** : mathematical formula for working out the average item intercorrelation in a test as an estimate of the average split-half correlations from all possible splittings of the test.
- logical construct** : a hypothetical quality or attribute such as intelligence, emotional stability and drive, that may explain unobserved events intervening between an independent variable (stimulus) and a behavioral response.
- norm** : representative value for a group of subjects in the scale of the transformed scores of a test.
- organismic variable** : such a physical or psychological characteristic of the subjects in a psychological experiment as may act as either an independent or an extraneous variable.
- percentile scale** : an ordinal scale formed by nonlinear transformation of raw test scores to percentile ranks.
- power test** : a test given enough time to enable not less than 75% of the subjects to attempt all its items which are arranged in an ascending order of difficulty.
- predictive validity** : suitability of a test in predicting a specific future performance, as given by the correlation between the test scores and the scores of a criterion measuring that subsequent performance.
- ratio IQ** : mental age of a child expressed as a percentage of his/her chronological age for the meaningful interpretation of the mental age score relative to the chronological age.
- relevant variable** : any extraneous variable which, though not deliberately used or intended to be used for studying its effect on the dependent variable, happens to affect the latter in an experiment.
- reliability** : consistency of results of a test on its repeated applications on the same sample under identical testing conditions.
- reliability coefficient** : the proportion of true variance in the total variance of the scores of a test, used as a measure of its reliability.
- speed test** : a test having test items of uniform difficulty levels and requiring the subjects to answer them within such a stipulated time as is too short for anyone to answer all the items.
- split-half reliability** : measure of equivalence and consistency of two split halves of a test, obtained by correlating the scores of the two halves administered to the same group of subjects either simultaneously or in immediate succession.
- standard scores** : either linearly or nonlinearly transformed raw scores such as z , T , C and stanine scores.
- stanine** : normalized standard scores, worked out by the nonlinear transformation of raw test scores, and forming a nine-point scale ranging from 1 to 9 with a mean of 5.0 and an SD of 1.96.
- stimulus variable** : any physical or social event that stimulates a specific behavioral response in the subjects of a test or experiment.
- T scale** : a scale of normalized standard scores, computed by the nonlinear transformation of raw test scores, and forming a distribution ranging from 0 to 100 with a mean of 50 and an SD of 10.
- test-retest reliability** : measure of reliability worked out by correlating two sets of scores of the same test applied twice on the same sample with a suitable intervening interval.
- validity** : capacity of a test to assess and predict a specific variable in exclusion of other closely related ones.
- validity coefficient** : measure of validity given by the correlation coefficient between the scores of the given test and those of an external criterion.

11. ANALYSIS OF VARIANCE

An experiment is designed to study the effects of *one or more independent variables* on a *single dependent variable* (pages 4-6) ; e.g., the effect of administered doses of a hypoglycemic agent (independent variable) on the blood sugar level (dependent variable) ; the effect of a practice schedule (independent variable) on the performance in a particular psychological test (dependent variable) ; the effect of doses of a pesticide (independent variable) on the tracheal ventilation (dependent variable) of locusts ; the effect of iodoacetamide (independent variable) on malt fermentation (dependent variable) by yeast.

Analysis of variance is used to find whether or not the exposure of the sample to the independent variable has enhanced the variance of the dependent variable significantly above its variance due to random factors.

11.1 EXPERIMENTAL DESIGN

Experimental design is the scientific planning of an experiment for exploring the effect of one or more chosen *independent variables* on a specific *dependent variable*. It includes the following procedures.

1. Opting for uncontrolled or controlled experiments :

In *uncontrolled investigations*, the sample has already been exposed to *uncontrolled classification variables* such as sex, age, race, habitat, genetic error, atmospheric factors, etc., as independent variables. The investigator cannot "fix" or control (i) the *levels* (i.e., amounts, amplitudes, intensities, doses, frequencies, etc.) of such independent variables, (ii) the *mechanism* of their application, and (iii) the *instants* and the *durations* of application; (iv) in some cases, the investigator *cannot*

apply the independent variable (e.g., a mutagen) directly on human subjects for studying its effects, and has to depend on human cases already exposed accidentally or in natural courses, or long before the experiment has been planned. Such independent variables are uncontrolled *classification variables* (page 5), free to *vary at random*. Use of such independent variables in an investigation leads to considerable *random errors* and ambiguous experimental observations. (v) Because of uncontrolled application of independent variables, the *relevant variables* (page 6) affecting the dependent variable cannot be effectively controlled or eliminated either; this adds further to the *experimental errors*. (vi) Uncontrolled experiments have often to be conducted by *retrospective methods* (see below) which are weaker, less precise and more prone to errors than prospective methods.

A few examples of uncontrolled experiments are : studying effects of adolescent tobacco smoking habits on adult emphysema or pulmonary carcinoma, of atmospheric sulfur dioxide on tracheal ventilation of beetles, of childhood measles infections on adult hepatic enzyme profile, of genetic mutations on metabolic pathways, of sex on serum LDL-cholesterol, or of atmospheric mercury vapours on pulmonary ventilation.

In *controlled investigations*, in contrast, independent variables are "*fixed*" *treatment variables* (page 5) whose levels, methods, modes and timings of application are rigorously determined and controlled by the investigator so as to eliminate random errors as far as possible. Moreover, because of controlled application, many of the *relevant variables* affecting the dependent variable may also be effectively contained or eliminated, thus

minimizing experimental errors. These enhance the precision of the experiment and lessen reasonably the ambiguities and errors in the inferences to be drawn. A few examples include the studies of effects of chosen doses of ofloxacin on intestinal bacterial flora profile, of selective prefrontal lobe ablations on the learning ability of anthropoid apes, of predetermined levels of a barbiturate on respiration of mitochondrial preparations, or of specific subtotal pancreatectomies on liver lipids. Generally, the more efficient and dependable prospective methods are used in investigating such controlled experiments.

2. Choosing retrospective or prospective methods :

Retrospective method is frequently used for uncontrolled experiments using *classification variables* beyond the control of the investigator (page 5). Here, the investigation is *carried backward in time* to explore a possible past cause of the events or changes already existing at the time of investigation. An *experimntal group* is constituted by a sample of individuals or cases showing specific changes in the dependent variable, seemingly due to past exposure to a postulated independent variable such as chronic smoking habits, alcohol or drug addiction, and prenatal infection. A *control group* is also drawn with individuals or cases not showing those changes in the dependent variable. The past history or the record of exposure to the postulated independent variable is then collected for individuals of both groups and compared for drawing inferences. For example, to find if prolonged standing (independent variable), past of venous varicosity (dependent variable), past activities involving prolonged standing are investigated in an experimental group of varicose patients and a control group of nonvaricose persons. Retrospective methods are useful in investigating (i) *long-term effects* such as the effects of prolonged smoking on

pulmonary emphysema, (ii) *rare events* like inherited diseases such as oculocutaneous albinism and Down syndrome, (iii) *random or classification type* of independent variables beyond the control of the investigator and suffering from *random errors*, and (iv) effects of such independent variables as *cannot be applied* on human samples directly (e.g., mutagens). Retrospective methods have following shortcomings : (i) past records may be incomplete and vague; (ii) samples drawn may not be unbiased, nor truly representative of the corresponding populations; (iii) restricted sampling from selected groups like indoor patients of a hospital may unduly exaggerate the frequency of rare events far above their proportion in the actual population; (iv) *relevant variables* affecting the dependent variable may escape from being eliminated and may thus lead to high experimental errors.

Prospective method is used mainly in *controlled experiments*, in which "*fixed*" *experimental treatments* are applied under the control of the investigator as independent variables (page 5). It is *preferred to the retrospective method*, except for long-term effects, rare events, and impracticability of applying the independent variable by the investigator himself. Here, the investigation is *continued forward in time* so as to explore the subsequent effects of the independent variable being applied now on a sample by the investigator himself. From a random sample, the investigator allocates individuals at random to the control and experimental groups. The individuals of the experimental group(s) are then exposed in a strictly regulated manner to chosen levels of a "*fixed*" treatment variable under his control; the individuals of the control group are excluded from the application of the independent variable. The dependent variable is subsequently measured in both experimental and control groups, and the findings of the groups are ompared. For example, an experi-

mental group of albino rats may be injected with the chosen level of thiouracil over a period while similar rats of a control group are injected with *placebo* containing only the solvent for the injectule with no drug; subsequently, the blood T_4 activity (dependent variable) is estimated in both the groups and compared for inferring whether thiouracil has significantly changed the blood T_4 activity.

3. Pilot experiment :

A pilot experiment using small groups should be carried out before further planning for the actual full-scale extensive investigation. It serves the following purposes; (i) it indicates the possibility of the *anticipated effect* of independent variable in the full-scale investigation to follow, (ii) it helps in deciding the levels of the independent variable to be applied on the sample and (iii) the number of replications of each level in the full-scale experiment to follow, (iv) it is used in working out the *sample size* for the full-scale experiment, and (v) it helps in planning for eliminating or minimizing the effects of numerous *relevant variables* likely to affect the dependent variable, and limits thereby the experimental errors.

4. Levels and replications :

Levels consist of the chosen amounts, intensities, amplitudes, categories, doses, etc., of the independent variable (factor) to be applied on the sample in the experiment. The number (k) of groups of individuals, used in an experiment, is determined by the number of levels of the factor to be applied. For example, the individuals of a random sample from the population would be randomly allocated to three groups ($k=3$) where three levels like 0, 5 and 10 micrograms of a hypoglycemic factor (independent variable) are planned to be administered to the subjects in different groups for studying the effect of that factor on the blood sugar (dependent variable) in an experiment. Each level of the factor would be

applied on the individuals of one of the groups while all the individuals of a group should be exposed to the same level of the factor.

To minimize experimental errors, each level of the factor (independent variable) should be applied on more than one individual, constituting a group. This is known as the *replication* of each level of the factor. The size (n) of each group, consisting of a number of individuals chosen at random from the sample and exposed to a given level of the factor, is determined by the desired number of replications of that level. For example, each of three groups should consist of 10 individuals ($n=10$) chosen at random from the sample, if ten replications of each of three levels of a factor are intended in an experiment; for this, a sample of 3×10 or 30 cases ($n \times k$) is to be drawn initially from the population.

5. Single-factor and factorial experiments :

A *single-factor or single-classification experiment* is undertaken to explore changes in the dependent variable due to the exposure of the groups from a sample to the levels of a *single independent variable*. Such an experiment may be designed in several ways as follows.

(a) *Independent group experiment* : For this, each group is constituted by a *separate set of individuals*, chosen at random from the sample drawn initially from the population depending on the laws of probability. None of the groups includes any individual, common to or associated with the other groups. The groups may be of identical or different sizes; *large independent groups* would consist of not less than 30 cases each ($n \geq 30$) while *small independent groups* would have group sizes lower than 30 ($n < 30$). Each such independent group would be exposed to a specific level of the independent variable so that there should be as many groups as the number of intended levels of the latter, while the group size would

depend upon the number of replications of the level to be applied on that group. One of these groups, called the *control group*, would be exposed to a level of treatment that is free from the independent variable; the others, called the *experimental groups*, would receive different other levels of treatment consisting of respective specific doses or amounts of the independent variable. As the groups consist of separate sets of independent individuals, the dependent variable scores of different groups would bear *no correlation* with each other (pages 124-125).

(b) *Single-group experiment* : In such an experiment, the same group of individuals, randomly sampled from a population, serves first as the *control group* and subsequently as the *experimental group(s)*; each time, the same group would be exposed to a separate specific level of the independent variable, after which the dependent variable scores would be measured in its individual cases (pages 132-133). Being constituted by the same single group of individuals, the control group as well as each experimental group has the *same size*. A *large single group* would consist of 30 or more individuals ($n \geq 30$) while a *small single group* would have less than 30 cases ($n < 30$). Because the same group of individuals is used for successive levels of exposure to the independent variable, each individual would yield a pair of dependent variable scores after every pair of consecutive treatment levels; the two scores of each such pair would bear a relation to one another. Thus, the dependent variable scores after successive treatment levels would constitute *paired and correlated observations*. For example, the kneejerk reflex strengths may be initially measured in a group of athletes after injecting them with placebo free from adrenaline (*control group*); the same group is next injected with a dose of adrenaline and the kneejerk strength is measured again (*experimental group*). These two sets of kneejerk scores form paired and correlated observations.

(c) *Matched-pair equivalent group experiment* : For this, either the intended dependent variable, or some other variable considered as related to the latter, is measured initially in all individuals of a random sample. Next, two *equivalent or matched-pair groups* are constituted by including in one group such individuals, each of whom is matched with an individual of the other group with respect to the initially measured variable (page 139). The two matched groups are then treated with two different levels of the independent variable; one of these groups may thus be treated with placebo (level 1) to serve as the *control group* while the other may be treated with a given dose of the independent variable (level 2) to serve as the *experimental group*. The dependent variable is subsequently measured in the individuals of both groups; these two sets of final scores of dependent variable constitute *paired and correlated observations*. Matched-pair groups are identical in size and may be large ($n \geq 30$) or small ($n < 30$).

(d) *Matched-mean equivalent group experiment* : Here, individuals are so allocated from sample to groups that group means of initially measured scores are equal, but individual scores of two groups may not be matched in pairs.

(e) *Randomized block experiment* : An experiment may be designed this way if the sample has been drawn at random from a population which has r number of classes, strata or categories bearing varied relations with the dependent variable to be studied. For this, the entire sample, consisting of n number of individuals, is divided into r number of *blocks*. Each block consists of k number of individuals belonging to a specific one of the classes or strata — k is also the number of *levels* of the independent variable planned to be applied. Thus, n equals $r \times k$. The individuals of each block are allocated at random to different levels of treatment so that (i) each level of treatment is applied on only one individual of each

block, and (ii) all the levels are applied on one or other of the members of each block. So, r is the number of *replications* of each of the k number of levels. After exposure of the blocks to all the levels of treatment, the dependent variable is finally measured in all the individuals of each block. The dependent variable scores of such individuals, as have been exposed to the same level of treatment, are then arranged in a group — this is repeated for all the levels of treatment used. The scores of different groups are then compared and analyzed statistically.

A *factorial experiment*, in contrast to the single-factor experiment, is designed to study the effect of combinations of different chosen levels of more than one independent variable (*factor*) on a given dependent variable. It may be a *two-way*, *three-way* or *four-way classification experiment* according to the number of factors applied. For example, in a two-way experiment to study the effects of cold exposure and thyrotropin administration on the thyroid size, the sample is divided into as many groups as the *number of combinations of levels* of the two independent variables and each group is made of as many individuals as the number of replications of the relevant combination. Thereafter, the individuals of each group are treated with a specific combination of levels of the independent variables.

6. Sample size :

The size of the sample should be worked out statistically before drawing a sample for the experiment. The sample should be sufficiently large to ensure that (i) the sample is truly representative of the population and includes individuals of different types or categories in the same proportions in which they occur in the population — the smaller the sample, the higher is the probability of non-inclusion of rare cases into it, thus making the sample less representative of the population, (ii) the distribution of scores of the dependent variable

in the sample conforms to either normal or t distributions so as to enable the use of parametric methods for statistical inferences from the experimental data, (iii) the *SD* of those scores does not suffer from downward bias due to the non-inclusion of rare cases from the population in the sample, (iv) a large sample size narrows the H_0 and H_a distributions to decrease the area of their overlap, and consequently lowers the probability of *type II error* (β) of inference (page 114) without enhancing the probability of *type I error* (α), (v) it thereby increases the *power* of a test in detecting genuine differences as its power is given by $(1 - \beta)$, and (vi) it decreases the errors due to planned exclusion of the rest of the population.

The sample size (n) is frequently estimated, using the unbiased *SD* (s) from the pilot experiment as an estimate for the population *SD* (σ), the *critical deviation* (D) which is the specified difference between the means (μ_0 and μ_a) of respectively the H_0 and H_a distributions, and a chosen level (α) of significance. For a *one-tail test* for difference between two means, either the critical t score with ∞ degrees of freedom and for the chosen one-tail α , or the critical z score (z_α) for that α , may be used for estimating n .

$$D = \mu_a - \mu_0; \quad n = \frac{s^2 z_\alpha^2}{D^2}, \quad \text{or } n = \frac{s^2 t_{\alpha(\infty)}^2}{D^2}.$$

For a *two-tail test*, either the critical $z_{\alpha/2}$ or the critical $t_{\alpha/2(\infty)}$ is used, beyond which lies the fractional area ($\alpha/2$) in each tail of respectively the normal and the t distributions.

$$n = \frac{s^2 z_{\alpha/2}^2}{D^2}, \quad \text{or } n = \frac{s^2 t_{\alpha/2(\infty)}^2}{D^2}.$$

7. Randomization :

Randomization should be ensured both in sampling and in treatment with the independent variable.

A. Random sampling :

To make a sample fairly representative of the population, *probability sampling* is used to choose at random the required number of individuals from a population for inclusion in the sample. This depends on (i) laws of probability and (ii) frequencies of different types of individuals in the population, and (iii) leaves no scope for arbitrary choice by any person.

(a) *Simple random sampling* : This is the random choice of individuals for a sample from the undivided whole of a *small, finite and homogeneous population*, depending on laws of probability. It ensures that (i) *no element of conscious or unconscious bias*, whim or personal factor of the investigator affects the choice, (ii) each member of the population has an *equal probability* of being chosen for the sample, (iii) the choice of any member is *independent* of the choice or exclusion of any other member, and (iv) *proportions of different types of individuals* in the sample conforms to their respective proportions in the population. In the *simple card drawal method*, all members of the population are given serial numbers which are entered on separate cards; after mixing the cards together, as many of them are picked up at random as the required sample size, and the corresponding individuals of those chosen cards are included in the sample. The more scientific *random number method* is based on the choices from digits of a random number table of such digits arranged in sequences with a random number generator (pages 8-9). For *sampling with replacement*, a member once chosen for the sample continues to be considered for all subsequent choices also, so as to keep the probability of choice of every individual unchanged and identical at any step of choice; but in *sampling without replacement*, an individual once chosen is kept out of all subsequent choices so that the probability of choice rises, though by small

degrees, from choice to choice instead of remaining unchanged — this needs a modification of formulae used subsequently in computing standard errors from such a sample (pages 74-75).

(b) *Stratified random sampling* : This method is used for a *large and heterogeneous population* with distinct strata. Here, simple random sampling is applied *separately on each stratum*, drawing as many individuals from the latter as would correspond to the proportional size of that stratum in the population; the proportions of individuals in different strata of the sample would thus conform to those in the respective strata of the population. Moreover, the probability of getting chosen remains *identical for all members of a stratum*, but *varies from stratum to stratum* according to the proportional sizes of strata in the population.

(c) *Multistage sampling* : This method is used for a *vast and widespread population*. Utilizing some existing steps in the population, it is divided into a number of successive stages ranging from the entire population to the individuals. Simple random sampling is applied at each stage separately and successively (page 9). For example, a few *first-stage units* are first randomly chosen from the population, then a few *second-stage units* are chosen at random from the chosen first-stage units, and a specific number of individuals may next be chosen from the chosen second-stage units, as the *third-stage units* to form the sample.

(d) *Fixed interval sampling* : Sometimes, all members of a population arrive, or turn out, or are *arranged serially in a systematic order*, such as the butterflies netted serially from air, fishes angled serially from a water-body, or patients arriving successively at the out-patient clinic. In such cases, the first member for the sample is chosen purely at random and is then followed by the choice of other members at

randomly predetermined regular intervals along the serial order of their turn-outs.

B. Random treatment :

Wherever possible, different levels of the fixed treatment variable should be applied in sequences varying at random from individual to individual of the sample. This would prevent or eliminate the *order effect* of the treatment, which may result if the levels of treatment are applied in the same sequence to all the individuals.

8. Minimizing effects of relevant variables :

In any experimental system, there always occur many *relevant variables* which are not intended to be applied in that experiment, but nevertheless may influence or affect the dependent variable of the experiment or otherwise sway the experimental result (page 6). Effects of all such relevant variables that may interfere with the aimed investigation need to be countered or minimized by proper designing of the experiment and its proceedings. Thus, *subject-relevant variables*, such as quantitative or qualitative variations of age, sex, race, strain, body weight, attitude, aptitude, genotype and phenotype of subjects chosen for the sample, may affect the experimental outcome, though not so desired, and should therefore be minimized by properly designed random probability sampling. *Situational relevant variables*, such as unintended variations of pH, temperature, ionic concentrations, osmolarity, noise decibels, light intensities, workshop ventilation, etc., in the experimental medium and system or in the surroundings, should be controlled and stabilized — by using buffers, thermostats, voltage stabilizer, etc., for instance — so as not to affect the investigation. *Sequence-relevant variables* should be eliminated by randomizing the sequence of applications of different levels of experimental treatments from subject to subject in the sample, to avoid the *order effect* of applications (see above).

9. Statistical treatments :

The data obtained in any experiment or investigation have eventually to be subjected to statistical tests for analysis, interpretation, inference and prediction. Because the application of each statistical test depends on the justifiability of its specific assumptions in the pertinent investigation, the latter should be so designed as to conform to the assumptions and justify the application of a powerful and parametric test to the obtained data. Where the significance of difference between group means would be sought after collecting the data, Wilcoxon's composite rank test, Student's *t* test and analysis of variance (anova) — three possible alternatives — possess progressively higher powers and dependabilities; so, the experimental design should endeavour to conform to the assumptions for anova and to enable thereby the use of that test in preference to the two preceding tests. If both a parametric test and its nonparametric alternative are available, the former should be preferable as it is more powerful than the latter so long as the assumptions for the former are justifiable; so, the experiment should be so designed as to aim at making the assumptions of the parametric test justifiable — the nonparametric test should be taken recourse to, only if the assumptions for its parametric alternative are not justifiable. This explains why the product-moment correlation (*r*) is preferable to Spearman's or Kendall's rank-dependent correlations, so long as the assumptions of *r* are justifiable; moreover, a significant *r*, in contrast to the other two, leads to further explorations by parametric partial and multiple linear correlations into associations involving more than the two originally explored variables (pages 162-167).

While designing an experiment, the qualitative and quantitative laboratory tests should also be chosen on the basis of their *reliabilities* and *validities*, as measured statis-

tically. An experiment or test has a higher reliability if it measures or assesses the intended variable more consistently with lesser errors, and possesses a higher validity if it can measure that variable in exclusion of most or all of the other closely related variables (pages 257 and 263).

11.2 ANALYSIS OF VARIANCE

Analysis of variance (*anova*) is an extension and generalization of Student's *t* test, but (i) is preferable to the latter because *anova* is far more powerful, (ii) can be applied to two or more groups simultaneously, (iii) can be used in estimating the strength of association between the dependent variable and the independent variable, and (iv) also helps in minimizing experimental errors because experiments have to be designed more rigorously to conform to its assumptions.

It tests the difference between the variances of two or more groups. For example, in the follow-up of an experiment using a single independent variable, the *anova* (one-way *anova*) analyzes different components of the total variance (s_t^2) of the sample to estimate the relative magnitudes of the within-groups variance (s_w^2) due to uncontrolled random factors, and the between-groups variance (s_b^2) which may have been influenced by the applied independent variable — it aims at finding out whether or not s_b^2 can be explained away by the null hypothesis that it does not differ significantly from the s_w^2 . The basic *anova* statistic is the *variance ratio* or *F* ratio (§ 11.3).

Classification of *anova*

The method of *anova* differs according to the *number of independent variable(s)* used in the experiment.

A *single-classification* or *one-way anova* is used to investigate the effects of a single

independent variable on the dependent variable. The number of applied levels of the independent variable determines the number of groups in the experiment. The size of each group equals the number of *replications* of the given level of the independent variable. For example, a one-way *anova* may be applied to tracheal ventilation values (dependent variable) measured in three groups of 20 insects each, after their exposure to three respective doses of a pesticide (independent variable), to find if the pesticide changes the tracheal ventilation significantly.

Higher orders of *anova* such as *two-way* and *three-way anovas* are used in a *factorial experiment* where the simultaneous effects of more than one independent variable are being investigated (page 285). The number of groups used corresponds to the chosen number of combinations of different levels of the independent variables, each such combination being applied on one of the groups. The size of each group corresponds to the desired number of replications of each combination of the independent variables.

Models of *anova*

Different models of *anova* have to be used according to the *nature(s) of independent variable(s)* in the experiment.

(a) *Fixed model* or *model I anova* :

This model is used in exploring “fixed” or controlled treatment effects. In other words, it analyzes the variances of a dependent variable in experiments using “fixed” *experimental treatment(s)* as independent variable(s). A model I *anova* is thus used for studying the effects of chosen and controlled levels of drugs, hormones, ions, radiations, experimental lesions or ablations of brain, temperature, pH, light, osmolality, etc., on physical properties, chemical constituents, structural components, activities, functional aspects and behaviours of organisms. It is also used in studying the

effects of practice, learning methods, etc., on performance. A one-way model I anova, for example, explores whether or not the between-groups variance (s_b^2) contains an *added treatment component*, owing to the exposure of the groups of subjects to different levels of a single treatment variable, distinct from and in addition to the variances associated with random sampling (§ 11.3). In model I anova, the independent variables, strictly controlled by the investigator, *do not suffer* from random changes. So, a model I anova, showing significant results, may be followed by the estimation of the *strength of association* between the dependent and the independent variables by working out *omega square* (§ 11.3).

(b) *Random model or model II anova :*

This model is used in exploring the effects of chosen random factors on the dependent variable. In other words, it analyzes the variances of a dependent variable in experiments where the groups have been exposed to independent variable(s) such as sex, race, age, genotypes, habitats, home environments, atmospheric temperature, atmospheric pollutants and cosmic rays, which are randomly changing *classification variables*, largely beyond the control of the investigator. Thus, it may explore the differences between groups drawn from different breeds, genotypes, phenotypes, natural habitats, races, socio-economic status and home environments. In a one-way model II anova, for example, it is investigated whether or not the between-groups variance (s_b^2) contains an *added variance component* (s_a^2), absent within the groups, but present between the latter subsequent to their exposure to the independent variable(s). In model II anova, the uncontrolled independent (classification) variables suffer from random fluctuations ; so, a significant result in model II anova is not followed by working out omega

square for the strength of association between the dependent and independent variables.

(c) *Mixed model or model III anova :*

Unlike models I and II which may be either one-way or of higher orders, model III is always a two-way or a still higher order of anova where some independent variable(s) must be "fixed" experimental treatment(s) while the other(s) must be uncontrolled classification variable(s). The effects of two or more levels of a "fixed" treatment variable may also be explored by model III anova in a *single-group experiment*, because the randomly chosen subjects or cases of the group constitute a classification variable affecting the dependent variable in addition to the treatment variable used (§ 11.6).

Assumptions of anova

(a) *Random assignment :*

(i) The experimental design should provide for *random sampling* so that each individual of the population has an equal probability of being chosen for a group, and the choice of each individual is independent of the choice of others.

(ii) *Randomization of treatment* should also be ensured for different levels of the independent variable(s), wherever possible. For this, all individuals of the sample should be allocated at random to different sequences of application of the chosen levels of the independent variable. The dependent variable is measured in each individual, following the treatment with each level of the independent variable. After all the levels have been applied on the individuals in sequences varying at random from individual to individual, the measured scores of the dependent variable are arranged into groups according to the respective applied levels of the independent variable. The scores of different groups, thus constituted, are then subjected to anova. Such

randomized treatment eliminates the *order effect* which may result if different levels are applied in the same sequence to all the individuals.

(b) *Normal distribution :*

The dependent variable should have a normal distribution in the population. Stated otherwise, it should be reasonable to assume that the *error terms*, i.e., the deviations of individual scores from the respective group means, are distributed normally.

(c) *Independence of errors :*

The *error terms*, i.e., the deviations of individual scores from the group mean, should be independent of each other. This is an alternative form of the assumption that the individual scores occur at random and independent of each other.

(d) *Homoscedasticity :*

The assumption of homoscedasticity implies that the groups drawn for an experiment possess homogeneous variances initially. In other words, they should have been drawn from the same population (or closely similar populations) so that their initial variances may be considered as different estimates of the same population variance, differing only due to their sampling errors. It should thus be reasonable to assume that the error terms of individuals of different groups have homogeneous dispersions.

(e) *Additivity :*

It should be reasonable to assume that different factors, including the independent variables used, produce separate bits of variations of the dependent variable and these variations add up to give the total variation of the latter. This additive property of variations, due to different factors, enables the analysis of the total variance (s_t^2) of the dependent variable into its various components.

Nonlinear transformations for non-normality of scores :

In case of a gross deviation of the observed scores of the dependent variable from the normal distribution, the following non-linear transformations of raw scores may be tried for getting normalized transformed scores (pages 75 and 265).

(i) *Logarithmic transformation :* This is tried for reducing the positive skewness of the raw score distribution by changing the raw scores into their logarithms ; where X_s is the transformed score, $X_s = \log X$. It is worked out when the variance of raw scores tends to be directly proportional to their squared mean. pH is such a transformation.

(ii) *Square root transformation :* It is often used to decrease the skewness of the raw score distribution where the variance of raw scores tends to be directly proportional to their mean. The transformed score is here the square root of the raw score : $X_s = \sqrt{X}$.

(iii) *Reciprocal transformation :* It is often undertaken when the *SD* of raw scores tends to be proportional to their squared mean. Each raw score is changed into its reciprocal : $X_s = 1/X$.

(iv) *Arcsine transformation :* It is tried when the variance of a distribution of proportions (p) is roughly proportional to $\bar{X}(1 - \bar{X})$. It serves to lengthen the tails of the distribution and to compress its central part. The transformed score is the angle (degree radian) whose sine is the square root of the raw proportion : $X_s = \sin^{-1} \sqrt{p}$.

The transformed scores in each of these cases are tested by either chi square or G test for the normality of their distribution. If the test indicates their significant goodness of fit with the normal distribution, the transformed scores may be used for anova. Otherwise, the

raw scores may be used for nonparametric anova (§ 11.5).

11.3 ONE-WAY ANOVA

A one-way anova investigates the effects of a single independent variable on the dependent variable. It is undertaken to find whether or not the exposure of different groups of subjects or cases to different levels of a single independent variable has produced significant differences in the variance between the groups. For the assumptions of one-way anova, see pages 280-281. One-way anova may be either *model I* or *model II* according as the independent variable is a "fixed" experimental treatment or an uncontrolled classification variable (pages 279-280); it cannot be a model III anova which is always a higher order of anova with more than one independent variable.

Computation of variance ratio

One-way anova consists of the computation and interpretation of the anova statistic F which is the *variance ratio* between the between-groups variance (s_b^2) and the within-groups variance (s_w^2) in this class of anova.

$$F = \frac{\text{between - groups variance}}{\text{within - groups variance}} = \frac{s_b^2}{s_w^2}.$$

So, for computing the F ratio in a one-way anova, the two variance components, viz., s_b^2 and s_w^2 , of the total variance (s_t^2) of the dependent variable have to be worked out. But s_t^2 cannot be directly partitioned into these components, because variances are not additive. However, the sums of squares as well as their degrees of freedom, the ratios of which are the respective variances, are additive. So, the total sum of squares is first partitioned into its component sums of squares which are then divided by their respective degrees of freedom to obtain the variance components.

1. Partitioning of sums of squares :

(a) The total sum of squares (SS_t) is the sum of squared deviations of all individual

scores (viz., X_1, X_2 , etc., of the respective groups) from their grand mean (\bar{X}). If there are k number of groups with the respective sizes of n_1, n_2, \dots, n_k ;

$$N = \sum n_i = n_1 + n_2 + \dots + n_k;$$

$$\bar{X} = \frac{\sum X_i}{N} = \frac{\sum X_1 + \sum X_2 + \dots + \sum X_k}{n_1 + n_2 + \dots + n_k};$$

$$SS_t = \sum \sum (X_i - \bar{X})^2 \\ = \sum (X_1 - \bar{X})^2 + \sum (X_2 - \bar{X})^2 + \dots + \sum (X_k - \bar{X})^2;$$

$$\text{or, } SS_t = \sum X_1^2 + \sum X_2^2 + \dots + \sum X_k^2 \\ - \frac{(\sum X_1 + \sum X_2 + \dots + \sum X_k)^2}{N}.$$

$$df_t = N - 1,$$

because one df is lost in using \bar{X} as an estimate of the parametric mean μ .

(b) The *between-groups sum of squares* (SS_b) is computed by multiplying the squared deviation of each group mean (\bar{X}_i) from the grand mean with the number (n_i) of scores in that group as its weight, and totalling the resulting products for all the groups. SS_b has $(k-1)$ degrees of freedom because k number of group means are being considered and one df is lost in using (\bar{X}) as an estimate of μ .

$$\bar{X}_1 = \frac{\sum X_1}{n_1}; \quad \bar{X}_2 = \frac{\sum X_2}{n_2}; \quad \dots \quad \bar{X}_k = \frac{\sum X_k}{n_k}.$$

$$SS_b = \sum [n_i (\bar{X}_i - \bar{X})^2] \\ = n_1 (\bar{X}_1 - \bar{X})^2 + n_2 (\bar{X}_2 - \bar{X})^2 + \dots \\ + n_k (\bar{X}_k - \bar{X})^2, \quad \text{or,}$$

$$SS_b = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \dots + \frac{(\sum X_k)^2}{n_k} \\ - \frac{(\sum X_1 + \sum X_2 + \dots + \sum X_k)^2}{N}.$$

$$df_b = k - 1.$$

(c) The *within-groups sum of squares* (SS_w) is the sum of squared deviations of all individual scores (X_i) from their respective group means (\bar{X}_i). It has $(N - k)$ as its *df* because there is a total of N number of scores, and k number of group means (\bar{X}_1, \bar{X}_2 , etc.) are used as estimates of μ .

$$\begin{aligned} SS_w &= \sum \sum (X_i - \bar{X}_i)^2 \\ &= \sum (X_1 - \bar{X}_1)^2 + \sum (X_2 - \bar{X}_2)^2 + \dots \\ &\quad + \sum (X_k - \bar{X}_k)^2. \end{aligned}$$

$$df_w = N - k.$$

However, SS_w is usually worked out as follows depending on the additive properties of the sum of squares and their *df*.

$$SS_w = SS_t - SS_b; \quad df_w = df_t - df_b = N - k.$$

2. Working out of variances and *F* ratio :

(a) Each sum of square is divided by its *df* to give the corresponding variance or *mean square* (s^2).

$$s_t^2 = \frac{SS_t}{df_t} = \frac{SS_t}{N-1}; \quad s_b^2 = \frac{SS_b}{df_b} = \frac{SS_b}{k-1};$$

$$s_w^2 = \frac{SS_w}{df_w} = \frac{SS_w}{N-k}.$$

Of these mean squares, s_t^2 is the *total variance*, serving as an estimate of the dispersions of scores of all the groups around the grand mean and having the *df* of $(N - 1)$. Similarly, s_w^2 is the *within-groups variance* or *error variance* ($df = N - k$) and constitutes an unbiased estimate of the population variance (σ^2), made on the basis of the deviations of individual scores of the groups from the respective group means. It results from the uncontrolled random variations of individual scores from the corresponding group means owing to random factors like genetic, environmental and social variables, and

estimates the variations produced by random sampling. On the contrary, s_b^2 is the *between-groups variance* — so long as all the groups may be justifiably assumed to have come from the same population, to obey homoscedasticity, and not to have changed significantly in spite of their exposure to different levels of the independent variable, s_b^2 serves as another independent estimate of the population variance (σ^2), made on the basis of the deviations of group means from the grand mean. In other words, so long as the variance of the dependent variable has not changed significantly owing to the independent variable, s_b^2 estimates the same error variance as is estimated by s_w^2 . But where the independent variable has brought about a significant change in the dependent variable, s_b^2 also includes an *added component* due to such change, over and above the error variance estimated in common by both s_b^2 and s_w^2 . As the variance due to the uncontrolled factors is common to both s_b^2 and s_w^2 , variances are not additive: $s_t^2 \neq s_b^2 + s_w^2$.

(b) The computed s_b^2 and s_w^2 are next used in working out the *F* ratio or *variance ratio*. The denominator of the *F* ratio is called the *denominator variance* or the *lesser mean square*, and is an estimate of the error variance of the dependent variable scores, resulting from unknown and uncontrolled factors associated with random sampling. The numerator of the *F* ratio is called the *numerator variance* or the *greater mean square*, and consists of the variance estimate whose source and significance are under investigation. The degrees of freedom of the *F* are those of the two variances used in its computation, viz., df_b of the greater mean square and df_w of the lesser mean square. Thus, for a one-way anova,

$$F = \frac{\text{greater mean square}}{\text{lesser mean square}} = \frac{s_b^2}{s_w^2}.$$

$$df \text{ of } F : k - 1, N - k.$$

3. Significance of computed F :

The H_0 for one-way anova contends that there is no significant difference between the groups and their means are estimates of the same population μ as estimated by the grand mean. If H_0 is correct, s_b^2 would not differ significantly from s_w^2 and would not contain any added component other than the error variance estimated by s_w^2 , so that the computed F (i.e., the s_b^2/s_w^2 ratio) would not differ significantly from 1.00. Thus, if the computed F is found not significant, there is no added component of variance in s_b^2 and no significant difference between the groups.

But where the computed F turns out to be significant, s_b^2 would contain a *significant added component* owing to variations produced between the groups by their exposure to different levels of the independent variable, and would thus be significantly higher than s_w^2 — the H_0 would then be rejected and it would be inferred that there are *significant differences* between the groups and between their means. The added component in the s_b^2 is called an *added treatment component* when due to the effect of a “fixed” treatment variable, and an *added variance component* when resulting from exposure to an uncontrolled classification variable as the independent variable.

For finding the probability P of correctness of the H_0 , the computed F is compared with

critical F values (F_α) for chosen levels of significance (α not exceeding 0.05). The *sampling distributions* of F values and so, their *probability distributions* also, depend for their forms on the df_b and df_w of the greater and lesser mean squares used in computing the F . F distributions are reverse J-shaped for very small degrees of freedom, but change into unimodal distributions with progressively declining skewness with the rise in the degrees of freedom. So, the appropriate F_α has to be found out in accordance with the df values of the computed F . Moreover, being a ratio of two variances, F cannot be negative ; so, the rejection region (α) of the F distribution is always the fractional area beyond the critical F_α in the right tail of the specific F distribution for the given df values (Fig. 11.1). If the computed F either equals or exceeds the critical F_α (Table H of Appendix), the probability P of correctness of the H_0 is considered too low ($P \leq \alpha$) and the H_0 cannot be retained — the computed F is then significant and it is inferred that there are significant differences between the group means. But if the computed F is lower than the critical F_α , the H_0 cannot be rejected ($P > \alpha$) — the computed F is not significant and there is no significant difference between the group means.

In one-way anova *between more than two groups*, a significant F ratio merely indicates significant differences between some or all of the groups, and must be followed by *multiple comparison tests* to find which specific groups differ significantly from each other ; some such cases of anova are cited in *Examples* under the next section (§ 11.4). But in one-way anova *between two groups only*, a significant F leads straightway to the inference that the two group means differ significantly, and so, needs no further follow-up by multiple comparison tests ; some such cases of anova are cited in *Examples 11.3.1-11.3.3*. However, in both types

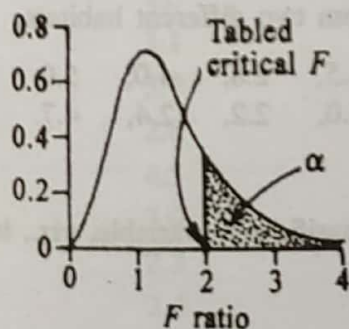


Fig. 11.1. Fractional area (α) beyond a critical F value in an F distribution.

of cases, a significant F ratio should be followed by the computation of either the *omega square* in case of model I anova, or the *added variance component* in case of model II anova.

On the contrary, if the computed F is not significant, neither any multiple comparison test nor the computation of omega square or added variance component is to follow.

Omega square

The independent variable in a model I anova is a "fixed" experimental treatment under the control of the investigator and so, does not suffer from random errors. So, if a model I anova yields a significant F , the *strength of association* between the independent variable and the dependent variable can be estimated by working out the omega square (ω^2). The value of the computed ω^2 gives that proportion of the total variance of the dependent variable which is associated with the independent variable.

$$\omega^2 = \frac{(k-1)(F-1)}{(k-1)(F-1) + N}$$

Added variance component

In model II anova, the uncontrolled

independent variable suffers from random errors and so, the strength of association between the independent variable and the dependent variable cannot be estimated here. So, instead of computing omega square, a significant F in model II anova is followed by working out an estimate of the *added variance component* (s_a^2) in the s_b^2 .

(a) Where all the groups are of equal size (n),

$$s_a^2 = \frac{s_b^2 - s_w^2}{n}$$

(b) But if there are k number of groups in the experiment and they differ in size from each other,

$$s_a^2 = \frac{s_b^2 - s_w^2}{\frac{1}{k-1} \left[(n_1 + n_2 + \dots + n_k) - \frac{n_1^2 + n_2^2 + \dots + n_k^2}{n_1 + n_2 + \dots + n_k} \right]}$$

In either case, proportionate variations are then worked out for s_a^2 and s_w^2 as follows :

$$(i) \text{ for } s_a^2 : \frac{s_a^2}{s_w^2 + s_a^2} ; \quad (ii) \text{ for } s_w^2 : \frac{s_w^2}{s_w^2 + s_a^2}$$

Example 11.3.1.

Apply one-way anova to find whether or not there is a significant difference between the mean winglengths (mm) of the following two groups of flies sampled from two different habitats.

Group 1	:	3.8,	4.3,	4.6,	5.1,	4.6,	5.3,	4.5,	2.8,	4.0,	5.0.
Group 2	:	3.0,	4.6,	3.1,	3.7,	2.8,	4.5,	3.0,	2.2,	2.4,	4.7.

Solution :

Because the independent variable consists of an uncontrolled classification variable, viz., habitat, a *one-way model II anova* is applicable here.

(a) *Partitioning of sums of squares :*

The data are entered in the first two columns of Table 11.1 and used in working out ΣX_1 , ΣX_2 , ΣX_1^2 and ΣX_2^2 .

Table 11.1. Table for working out sums of squares directly from raw winglength scores.

Group 1 (X_1)	Group 2 (X_2)	X_1^2	X_2^2
3.8	3.0	14.44	9.00
4.3	4.6	18.49	21.16
4.6	3.1	21.16	9.61
5.1	3.7	26.01	13.69
4.6	2.8	21.16	7.84
5.3	4.5	28.09	20.25
4.5	3.0	20.25	9.00
2.8	2.2	7.84	4.84
4.0	2.4	16.00	5.76
5.0	4.7	25.00	22.09
Total	44.0 (ΣX_1)	198.44 (ΣX_1^2)	123.24 (ΣX_2^2)

$$n_1 = 10 ; \quad n_2 = 10 ; \quad N = n_1 + n_2 = 10 + 10 = 20.$$

$$SS_t = \Sigma X_1^2 + \Sigma X_2^2 - \frac{(\Sigma X_1 + \Sigma X_2)^2}{N} = 198.44 + 123.24 - \frac{(44.0 + 34.0)^2}{20} = 17.48.$$

$$df_t = N - 1 = 20 - 1 = 19.$$

$$SS_b = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} - \frac{(\Sigma X_1 + \Sigma X_2)^2}{N} = \frac{(44.0)^2}{10} + \frac{(34.0)^2}{10} - \frac{(44.0 + 34.0)^2}{20} = 5.00.$$

$$df_b = k - 1 = 2 - 1 = 1.$$

$$SS_w = SS_t - SS_b = 17.48 - 5.00 = 12.48.$$

$$df_w = N - k = 20 - 2 = 18.$$

Alternatively, the data are entered in the first two columns of Table 11.2 and used in working out ΣX_1 , ΣX_2 , $\Sigma(X_1 - \bar{X})^2$ and $\Sigma(X_2 - \bar{X})^2$.

Table 11.2. Table for working out sums of squares using the means of winglength scores.

X_1	X_2	$X_1 - \bar{X}$	$(X_1 - \bar{X})^2$	$X_2 - \bar{X}$	$(X_2 - \bar{X})^2$
3.8	3.0	-0.1	0.01	-0.9	0.81
4.3	4.6	0.4	0.16	0.7	0.49
4.6	3.1	0.7	0.49	-0.8	0.64
5.1	3.7	1.2	1.44	-0.2	0.04
4.6	2.8	0.7	0.49	-1.1	1.21
5.3	4.5	1.4	1.96	0.6	0.36
4.5	3.0	0.6	0.36	-0.9	0.81
2.8	2.2	-1.1	1.21	-1.7	2.89
4.0	2.4	0.1	0.01	-1.5	2.25
5.0	4.7	1.1	1.21	0.8	0.64
Σ 44.0	34.0		7.34		10.14

$$n_1 = 10 ; \quad n_2 = 10 ; \quad N = n_1 + n_2 = 10 + 10 = 20.$$

$$\bar{X}_1 = \frac{\sum X_1}{n_1} = \frac{44.0}{10} = 4.4 ; \quad \bar{X}_2 = \frac{\sum X_2}{n_2} = \frac{34.0}{10} = 3.4. \quad \bar{X} = \frac{\sum X_1 + \sum X_2}{N} = \frac{44.0 + 34.0}{20} = 3.9.$$

$$SS_t = \sum (X_1 - \bar{X})^2 + \sum (X_2 - \bar{X})^2 = 7.34 + 10.14 = 17.48. \quad df_t = N - 1 = 20 - 1 = 19.$$

$$SS_b = n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 = 10(4.4 - 3.9)^2 + 10(3.4 - 3.9)^2 = 5.00.$$

$$df_b = k - 1 = 2 - 1 = 1.$$

$$SS_w = SS_t - SS_b = 17.48 - 5.00 = 12.48.$$

$$df_w = N - k = 20 - 2 = 18.$$

(b) Computation of variances and F ratio :

$$s_b^2 = \frac{SS_b}{df_b} = \frac{5.00}{1} = 5.00. \quad s_w^2 = \frac{SS_w}{df_w} = \frac{12.48}{18} = 0.69. \quad F = \frac{s_b^2}{s_w^2} = \frac{5.00}{0.69} = 7.25.$$

Table 11.3. Anova table for winglength data.

Sources of variation	Sums of squares	df	Variances	F
Between groups	5.00	1	5.00	7.25
Within groups	12.48	18	0.69	
Total	17.48	19		

(c) Significance of F :

Critical F values ($df_b = 1$, $df_w = 18$) are quoted from Table H of Appendix, for 2-tail test.

$$F_{.05(1,18)} = 4.41 ; \quad F_{.01(1,18)} = 8.28.$$

As the computed F exceeds the critical F for 0.05 level of significance, the probability P of the H_0 being correct is considered too low ; so, it is inferred that there is a *significant added variance component* between the groups, and that the two group means *differ significantly* ($P < 0.05$).

(d) Added variance component :

Size of each group (n) = 10.

$$s_a^2 = \frac{s_b^2 - s_w^2}{n} = \frac{5.00 - 0.69}{10} = 0.43. \quad \frac{s_a^2}{s_w^2 + s_a^2} = \frac{0.43}{0.69 + 0.43} = 0.38.$$

Example 11.3.2.

Apply one-way anova to find whether or not there is a significant difference between the mean tracheal ventilation scores (ml/min) of the following two groups of beetles treated respectively with two different doses of a pesticide.

Group 1	:	80,	81,	75,	80,	88,	70,	74,	71,	84,	72.
Group 2	:	70,	74,	68,	67,	72,	59,	61,	57,	68,	54.

Solution :

Because the independent variable consists of a "fixed" treatment variable in the form of controlled doses of a pesticide, a *one-way model I anova* is applicable here.

(a) *Partitioning of sums of squares :*

The data are entered in the first two columns of Table 11.4 and used for working out ΣX_1 , ΣX_2 , $\Sigma(X_1 - \bar{X})^2$ and $\Sigma(X_2 - \bar{X})^2$.

$$n_1 = 10 ; \quad n_2 = 10, \quad N = n_1 + n_2 = 10 + 10 = 20.$$

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{775}{10} = 77.50 ; \quad \bar{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{650}{10} = 65.00 ;$$

$$\bar{X} = \frac{\Sigma X_1 + \Sigma X_2}{N} = \frac{775 + 650}{20} = 71.25.$$

$$SS_t = \Sigma(X_1 - \bar{X})^2 + \Sigma(X_2 - \bar{X})^2 = 715.125 + 804.625 = 1519.75.$$

$$df_t = N - 1 = 20 - 1 = 19.$$

$$SS_b = n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 = 10(77.50 - 71.25)^2 + 10(65.00 - 71.25)^2 = 781.25.$$

$$df_b = k - 1 = 2 - 1 = 1.$$

$$SS_w = SS_t - SS_b = 1519.75 - 781.25 = 738.50. \quad df_w = N - k = 20 - 2 = 18.$$

Table 11.4. Table for computing sums of squares using the means of tracheal ventilation scores.

Group 1 (X_1)	Group 2 (X_2)	$X_1 - \bar{X}$	$(X_1 - \bar{X})^2$	$X_2 - \bar{X}$	$(X_2 - \bar{X})^2$
80	70	8.75	76.5625	- 1.25	1.5625
81	74	9.75	95.0625	2.75	7.5625
75	68	3.75	14.0625	- 3.25	10.5625
80	67	8.75	76.5625	- 4.25	18.0625
88	72	16.75	280.5625	0.75	0.5625
70	59	- 1.25	1.5625	- 12.25	150.0625
74	61	2.75	7.5625	- 10.25	105.0625
71	57	- 0.25	0.0625	- 14.25	203.0625
84	68	12.75	162.5625	- 3.25	10.5625
72	54	0.75	0.5625	- 17.25	297.5625
Σ 775	650		715.1250		804.6250

Alternatively, the data are entered in the first two columns of Table 11.5 and used in working out ΣX_1 , ΣX_2 , ΣX_1^2 and ΣX_2^2 .

Table 11.5. Table for computing sums of squares directly from raw tracheal ventilation scores.

X_1	X_2	X_1^2	X_2^2
80	70	6400	4900
81	74	6561	5476
75	68	5625	4624
80	67	6400	4489
88	72	7744	5184
70	59	4900	3481
74	61	5476	3721
71	57	5041	3249
84	68	7056	4624
72	54	5184	2916
Σ 775	650	60387	42664

$$n_1 = 10 ; \quad n_2 = 10 ; \quad N = n_1 + n_2 = 10 + 10 = 20.$$

$$SS_t = \Sigma X_1^2 + \Sigma X_2^2 - \frac{(\Sigma X_1 + \Sigma X_2)^2}{N} = 60387 + 42664 - \frac{(775 + 650)^2}{20} = 1519.75.$$

$$df_t = N - 1 = 20 - 1 = 19.$$

$$SS_b = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} - \frac{(\Sigma X_1 + \Sigma X_2)^2}{N} = \frac{(775)^2}{10} + \frac{(650)^2}{10} - \frac{(775 + 650)^2}{20} = 781.25$$

$$df_b = k - 1 = 2 - 1 = 1.$$

$$SS_w = SS_t - SS_b = 1519.75 - 781.25 = 738.50. \quad df_w = N - k = 20 - 2 = 18.$$

(b) Computation of variances and F ratio :

$$s_b^2 = \frac{SS_b}{df_b} = \frac{781.25}{1} = 781.25. \quad s_w^2 = \frac{SS_w}{df_w} = \frac{738.50}{18} = 41.03. \quad F = \frac{s_b^2}{s_w^2} = \frac{781.25}{41.03} = 19.04.$$

Table 11.6. Anova table for tracheal ventilation data.

Sources of variation	Sums of squares	df	Variances	F
Between groups	781.25	1		
Within groups	738.50	18	781.25	19.04
Total	1519.75	19	41.03	

(c) Significance of F :

Critical F values ($df_b = 1$, $df_w = 18$) are quoted from Table H of Appendix, for two-tail test.

$$F_{.05(1,18)} = 4.41 ; \quad F_{.01(1,18)} = 8.28.$$

As the computed F exceeds even the critical F for 0.01 level, the probability P for correctness of the H_0 is considered too low ; so, it is inferred that there is a significant added treatment component between the groups, and the two groups differ significantly ($P < 0.01$).

(d) *Omega square* :

As the model I anova has a significant F ratio, the strength of association is estimated between the independent variable and the dependent variable by working out the omega square.

$$\omega^2 = \frac{(k-1)(F-1)}{(k-1)(F-1) + N} = \frac{(2-1)(19.04-1)}{(2-1)(19.04-1) + 20} = 0.47.$$

So, a proportion of 0.47 of the total variance of tracheal ventilation is associated with the pesticide used as the treatment variable.

Example 11.3.3.

Following are the performance test scores of two groups of students after they have practised according to two respective practice schedules given to them. Do the groups differ significantly in performance due to their different practice schedules ?

Group 1	:	13,	20,	21,	20,	18,	12,	15,	23,	17,	13.
Group 2	:	28,	16,	30,	24,	20,	14,	32,	30,	20,	27, 28.

Solution :

A one-way model I anova is appropriate here because the two practice schedules constitute a controlled ("fixed") treatment variable.

(a) *Partitioning of sums of squares :*

The data are entered in the first two columns of Table 11.7 and used in working out ΣX_1 , ΣX_2 , ΣX_1^2 and ΣX_2^2 .

Table 11.7. Table for computing sums of squares directly from raw scores of performance test.

Group 1 (X_1)	Group 2 (X_2)	X_1^2	X_2^2
13	28	169	784
20	16	400	256
21	30	441	900
20	24	400	576
18	20	324	400
12	14	144	196
15	32	225	1024
23	30	529	900
17	20	289	400
13	27	169	729
	28		784
Total 172	269	3090	6949

$$n_1 = 10 ; \quad n_2 = 11 ; \quad N = n_1 + n_2 = 10 + 11 = 21.$$

$$SS_t = \sum X_1^2 + \sum X_2^2 - \frac{(\sum X_1 + \sum X_2)^2}{N} = 3090 + 6949 - \frac{(172 + 269)^2}{21} = 778.00,$$

$$df_t = N - 1 = 21 - 1 = 20.$$

$$SS_b = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} - \frac{(\sum X_1 + \sum X_2)^2}{N} = \frac{(172)^2}{10} + \frac{(269)^2}{11} - \frac{(172 + 269)^2}{21} = 275.67,$$

$$df_b = k - 1 = 2 - 1 = 1.$$

$$SS_w = SS_t - SS_b = 778.00 - 275.67 = 502.33. \quad df_w = N - k = 21 - 2 = 19.$$

(b) Computation of variances and F ratio :

$$s_b^2 = \frac{SS_b}{df_b} = \frac{275.67}{1} = 275.67. \quad s_w^2 = \frac{SS_w}{df_w} = \frac{502.33}{19} = 26.44. \quad F = \frac{s_b^2}{s_w^2} = \frac{275.67}{26.44} = 10.43.$$

Table 11.8. Anova table for performance score data.

Sources of variation	Sums of squares	df	Variances	F
Between groups	275.67	1	275.67	10.43
Within groups	502.33	19	26.44	
Total	778.00	20		

(c) Significance of F :

Two-tail critical F values ($df_b = 1, df_w = 19$) are quoted from Table H of Appendix.

$$F_{.05(1,19)} = 4.38; \quad F_{.01(1,19)} = 8.18.$$

As the computed F exceeds even the critical F for 0.01 level, the probability P of the H_0 being correct is considered too low. So, it is inferred that there is a significant added treatment component between the groups and the groups differ significantly ($P < 0.01$).

(d) Omega square :

As the model I anova has a significant F ratio, omega square is worked out to estimate the strength of association between the independent variable and the dependent variable.

$$\omega^2 = \frac{(k-1)(F-1)}{(k-1)(F-1) + N} = \frac{(2-1)(10.43-1)}{(2-1)(10.43-1) + 21} \approx 0.31.$$

So, a proportion of 0.31 of the total variance of the performance test scores is associated with the practice schedule used as the treatment variable.

11.4 MULTIPLE COMPARISON TESTS

If the F ratio is found to be significant in an anova with more than two groups, it should be followed by a multiple comparison test to find which group means differ significantly from each other. There are two types of multiple comparison tests.

Before-design or a-priori comparisons :

In some cases, even before commencing the experiment and collecting the data, the investigator can plan which group means should be subjected to a test for finding the significance of their differences. Statistical tests for such pre-planned comparisons of chosen

group means are called *a-priori* or *before-design comparisons*; e.g., multiple comparison *t* test and Scheffe's *F* test. The choice of group means to be compared, and of the test to be applied, depends on the purpose and design of the experiment in such cases, not on the examination and scrutiny of the data after the experiment is over.

After-design or *a-posteriori* comparisons :

In some cases, however, nothing can be so planned initially about which group means should be finally compared. After completion of the experiment in such cases, the data are subjected to a preliminary *F* test (§ 11.3) to find whether there is any significant difference at all between the group means of the data. If the *F* ratio is found to be significant, an *after-design* or *a-posteriori comparison* like Gabriel's sum of squares simultaneous test procedure (SS-STP) may be applied to explore the group means for significant differences.

Multiple comparison *t* test

Student's *t* is computed from the difference between the means of each pair of groups chosen *before* the experiment has been actually performed. Thus, this is a *before-design* or *a-priori comparison*. If the anova has yielded significant *F* ratio, one multiple comparison *t* test has to be undertaken for each chosen pair of group means. For each such test, the difference between the group means of the chosen pair, say $(\bar{X}_1 - \bar{X}_2)$, is converted to *t* score, using the *SE* of their difference ($s_{\bar{X}_1 - \bar{X}_2}$) — the latter is computed using the within-groups variance or error variance (s_w^2) used earlier in calculating the *F* ratio. Thus, for the difference $(\bar{X}_1 - \bar{X}_2)$,

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_2}} ;$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_2}}} ; \quad df = N - k ;$$

where n_1 and n_2 are the sizes of the respective groups, N is the total of all the group sizes in the experiment, k is the number of groups in the experiment, and $(N - k)$ is the df_w of the within-groups variance or lesser mean square.

If the computed *t* exceeds or equals the critical *t* score ($df = N - k$) for the chosen level of significance (α), the relevant group means are considered to differ significantly ($P \leq \alpha$). This procedure is repeated for group means of each chosen pair, using s_w^2 and the *sizes of those two groups*.

t test with Bonferroni adjustment

In the Bonferroni method, the difference between the means of each chosen pair of groups is converted as above to *t* score, using s_w^2 and the group sizes. The probability level (P) for each such computed *t* score is found out as above, by comparing the latter with the critical *t* scores having the same *df*. This P is then multiplied by the total number (k') of pairs of groups, being compared, to find the adjusted probability (P'). If this P' is lower than or equal to either a chosen α or the α of 0.05, relevant group means are considered to differ significantly from each other at or beyond that α ($P' \leq \alpha$).

Scheffe's *F* test

It is a powerful *a-priori* multiple comparison test, applicable inspite of minor degrees of skewness and heterogeneity of the distribution of scores, and inequality of the group sizes. It restricts type I errors of inference considerably, but enhances the type II errors.

For this test, *F* is computed from the difference between the means of each pair of

groups, chosen prior to the experiment, using s_w^2 and the sizes of the chosen groups. Thus, for the difference $(\bar{X}_1 - \bar{X}_2)$,

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_2}};$$

$$F = \frac{(\bar{X}_1 - \bar{X}_2)^2}{s_{\bar{X}_1 - \bar{X}_2}^2} = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_2}}.$$

Each computed F is then compared with critical F' values for 0.05 and 0.01 levels of significance. Each critical F' has to be computed by multiplying $(k - 1)$ with the critical F (df_b, df_w) for the given level of α , taken from Table H of Appendix, k being the total number of groups being compared. For example, where only two of the group means are being compared, and s_b^2 and s_w^2 have respectively 4 and 26 as their df , critical F' values for 0.05 and 0.01 levels are given by :

$$F'_{.05} = (k - 1) F_{.05(4,26)} = (2 - 1) \times 2.74 = 2.74;$$

$$F'_{.01} = (k - 1) F_{.01(4,26)} = (2 - 1) \times 4.14 = 4.14.$$

If the computed F of any pair of group means exceeds or equals the critical F'_α value, the difference between those group means is significant at or beyond that α ($P \leq \alpha$). This procedure is repeated for all other chosen pairs of group means.

Gabriel's SS-STP

The sum of squares simultaneous test procedure is used as an *a-posteriori* test.

(a) First, the critical sum of squares (SS_α) is computed, using the critical F value (F_α) for the chosen significance level (α) and with the degrees of freedom (df_b and df_w) of the greater and lesser mean squares respectively. Where k is the total number of groups, and s_w^2 is the within-groups variance,

$$SS_\alpha = (k - 1) s_w^2 F_{\alpha(df_b, df_w)}.$$

(b) The group means are computed from the obtained data and scrutinized to identify those group means which apparently look so close as to raise doubts about any significant difference between them. The *between-groups sum of squares* (SS_b) is computed between those groups. If, for example, the means of groups 2, 3 and 5 appear to be too close in the data obtained in an experiment, SS_b is computed between them, using the sums of scores ($\Sigma X_2, \Sigma X_3, \Sigma X_5$) of the respective groups and the respective group sizes (n_2, n_3, n_5).

$$SS_b = \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} + \frac{(\Sigma X_5)^2}{n_5} - \frac{(\Sigma X_2 + \Sigma X_3 + \Sigma X_5)^2}{n_2 + n_3 + n_5}.$$

(c) The computed SS_b is next compared with the computed SS_α . There is a significant difference between the chosen group means (\bar{X}_2, \bar{X}_3 and \bar{X}_5 in this case) only if the computed SS_b equals or exceeds the SS_α ($P \leq \alpha$).

(d) The data are next inspected to find if any other group mean apparently differs considerably from the group means tested in the preceding steps. SS_b is then computed with any such group mean (say, group 4 in the example cited).

$$SS_b = \frac{(\Sigma X_4)^2}{n_4} + \frac{(\Sigma X_2 + \Sigma X_3 + \Sigma X_5)^2}{n_2 + n_3 + n_5} - \frac{(\Sigma X_4 + \Sigma X_2 + \Sigma X_3 + \Sigma X_5)^2}{n_4 + n_2 + n_3 + n_5}.$$

(e) This computed SS_b is next compared with the SS_α computed as in step (a). Only if this SS_b equals or exceeds the SS_α , the group mean being tested differs significantly from the other group means tested in steps (b) and (c) above.

Example 11.4.1.

Blood sugar (mg dL⁻¹) was estimated in three groups of animals one hour after injecting the groups 1, 2 and 3 with respectively the placebo, 100 μ g of an anti-diabetic substance and 150 μ g of the latter. The blood sugar scores are recorded in the first three columns of Table 11.9. Is there any significant difference between the means of groups 1 and 2, and between those of groups 2 and 3? Also, estimate the strength of association between blood sugar and the anti-diabetic substance, if F ratio is significant.

Solution :

Because the independent variable consists of "fixed" treatments with different levels of the anti-diabetic substance, a *one-way model I anova* is applicable here.

(a) *Partitioning of sums of squares :*

The data entered in the first three columns of Table 11.9 are used in working out ΣX_1 , ΣX_2 , ΣX_3 , ΣX_1^2 , ΣX_2^2 and ΣX_3^2 .

Table 11.9. Table for computing sums of squares of blood sugar data.

Blood sugar scores			X_1^2	X_2^2	X_3^2
X_1	X_2	X_3			
128	118	88	16384	13924	7744
124	115	80	15376	13225	6400
129	122	84	16641	14884	7056
135	128	96	18225	16384	9216
132	125	86	17424	15625	7396
128	117	87	16384	13689	7569
118	110	96	13924	12100	9216
123	116	78	15129	13456	6084
117	108	98	13689	11664	9604
133	126	99	17689	15876	9801
Σ 1267	1185	892	160865	140827	80086

$$n_1 = 10 ; \quad n_2 = 10 ; \quad n_3 = 10 ; \quad N = n_1 + n_2 + n_3 = 10 + 10 + 10 = 30.$$

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{1267}{10} = 126.7 \text{ mg} ; \quad \bar{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{1185}{10} = 118.5 \text{ mg} ; \quad \bar{X}_3 = \frac{\Sigma X_3}{n_3} = \frac{892}{10} = 89.2 \text{ mg}.$$

$$\begin{aligned} SS_t &= \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 - \frac{(\Sigma X_1 + \Sigma X_2 + \Sigma X_3)^2}{N} \\ &= 160865 + 140827 + 80086 - \frac{(1267 + 1185 + 892)^2}{30} = 9033.47. \end{aligned}$$

$$df_t = N - 1 = 30 - 1 = 29.$$

$$\begin{aligned} SS_b &= \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} - \frac{(\Sigma X_1 + \Sigma X_2 + \Sigma X_3)^2}{N} \\ &= \frac{(1267)^2}{10} + \frac{(1185)^2}{10} + \frac{(892)^2}{10} - \frac{(1267 + 1185 + 892)^2}{30} = 7773.27 \end{aligned}$$

$$df_b = k - 1 = 3 - 1 = 2.$$

$$SS_w = SS_t - SS_b = 9033.47 - 7773.27 = 1260.20$$

$$df_w = N - k = 30 - 3 = 27.$$

(b) Computation of variances and F ratio :

$$s_b^2 = \frac{SS_b}{df_b} = \frac{7773.27}{2} = 3886.64 ; \quad s_w^2 = \frac{SS_w}{df_w} = \frac{1260.20}{27} = 46.67 ; \quad F = \frac{s_b^2}{s_w^2} = \frac{3886.64}{46.67} = 83.28.$$

df of computed F : $df_b, df_w = 2, 27$.

Table 11.10. Anova table for blood sugar data.

Sources of variation	Sums of squares	df	Variances	F
Between groups	7773.27	2	3886.64	83.28
Within groups	1260.20	27	46.67	
Total	9033.47	29		

(c) Significance of F :

Critical F values ($df = 2, 27$) are quoted below from Table H of Appendix.

$$F_{.05(2, 27)} = 3.35 ; \quad F_{.01(2, 27)} = 5.49.$$

As the computed F is found to be higher than the critical F for 0.01 level, the computed F is significant beyond the 0.01 level ($P < 0.01$). Hence, there is a *significant added treatment component* between the groups.

(d) Multiple comparison tests :

Scheffe's multiple comparison F test may be done to find out whether $(\bar{X}_1 - \bar{X}_2)$ and $(\bar{X}_2 - \bar{X}_3)$ are significant. As three group means are being compared, $k = 3$.

$$F = \frac{(\bar{X}_1 - \bar{X}_2)^2}{s_w^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \frac{(126.7 - 118.5)^2}{46.67 \left(\frac{1}{10} + \frac{1}{10} \right)} = 7.20 ; \quad F = \frac{(\bar{X}_2 - \bar{X}_3)^2}{s_w^2 \left(\frac{1}{n_2} + \frac{1}{n_3} \right)} = \frac{(118.5 - 89.2)^2}{46.67 \left(\frac{1}{10} + \frac{1}{10} \right)} = 91.97.$$

$$\text{Critical } F'_{.01} = (k - 1) F_{.01(2, 27)} = (3 - 1) \times 5.49 = 10.98.$$

$$\text{Critical } F'_{.05} = (k - 1) F_{.05(2, 27)} = (3 - 1) \times 3.35 = 6.70.$$

As the F ratio for $(\bar{X}_1 - \bar{X}_2)$ exceeds the critical $F'_{.05}$ but not the critical $F'_{.01}$, the difference between \bar{X}_1 and \bar{X}_2 is significant beyond the 0.05 level only ($P < 0.05$). But as the F ratio for $(\bar{X}_2 - \bar{X}_3)$ exceeds also the critical $F'_{.01}$, the difference between \bar{X}_2 and \bar{X}_3 is significant beyond the 0.01 level ($P < 0.01$).

[Alternatively, multiple comparison t test may be worked out with Bonferroni modification.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_2}}} = \frac{126.7 - 118.5}{\sqrt{\frac{46.67}{10} + \frac{46.67}{10}}} = 2.684 ; \quad t = \frac{\bar{X}_2 - \bar{X}_3}{\sqrt{\frac{s_w^2}{n_2} + \frac{s_w^2}{n_3}}} = \frac{118.5 - 89.2}{\sqrt{\frac{46.67}{10} + \frac{46.67}{10}}} = 9.590.$$

$$df = N - k = 30 - 3 = 27.$$

Critical t scores from Table B of Appendix :

$$t_{.02(27)} = 2.473 ; \quad t_{.01(27)} = 2.771 ; \quad t_{.001(27)} = 3.690.$$

Comparing the computed t scores with the critical t scores, it is found that $(\bar{X}_1 - \bar{X}_2)$ is significant beyond the 0.02 level ($P < 0.02$) while $(\bar{X}_2 - \bar{X}_3)$ is significant beyond the 0.001 level ($P < 0.001$).

To apply the *Bonferroni modification*, each P obtained by the t test is multiplied by the number k' of paired comparisons to get the corrected probability P' of the H_0 being correct.

$$k' = 2 ; \quad P' = k'P ;$$

$$\therefore \text{For } (\bar{X}_1 - \bar{X}_2) : P' = k'P = 2 \times (< 0.02) = < 0.04 ;$$

$$\text{for } (\bar{X}_2 - \bar{X}_3) : P' = k'P = 2 \times (< 0.001) = < 0.002.$$

Thus, $(\bar{X}_1 - \bar{X}_2)$ is significant beyond the 0.04 level ($P < 0.04$) while $(\bar{X}_2 - \bar{X}_3)$ is significant beyond the 0.002 level ($P < 0.002$).]

(e) *Strength of association :*

Omega square is computed to estimate the strength of association between blood sugar and the anti-diabetic factor. Using the F ratio computed in the model I anova and the number k of groups,

$$\omega^2 = \frac{(k-1)(F-1)}{(k-1)(F-1) + N} = \frac{(3-1)(83.28-1)}{(3-1)(83.28-1) + 30} = 0.85.$$

So, a proportion of 0.85 of the total variance of blood sugar is associated with the anti-diabetic factor used as the treatment variable.

Example 11.4.2.

The scores obtained in an abstract reasoning test by three groups of students, drawn from three different socioeconomic strata, are given in the first three columns of Table 11.11. Apply anova to find whether an added variance component is present in the variance between the groups due to random socioeconomic factors ($\alpha = 0.01$).

Solution :

The data, arranged in Table 11.11, are subjected to a *one-way model II anova* because the socioeconomic strata constitute a classification variable beyond the control of the investigator.

(a) *Partitioning of sums of squares :*

The data entered in the first three columns of Table 11.11 are used in working out the group means and the sums of squares.

Table 11.11. Table computing sums of squares from means of abstract reasoning scores.

X_1	X_2	X_3	$X_1 - \bar{X}$	$(X_1 - \bar{X})^2$	$X_2 - \bar{X}$	$(X_2 - \bar{X})^2$	$X_3 - \bar{X}$	$(X_3 - \bar{X})^2$
26	30	34	- 4.4	19.36	- 0.4	0.16	+ 3.6	12.96
27	34	35	- 3.4	11.56	+ 3.6	12.96	+ 4.6	21.16
25	28	28	- 5.4	29.16	- 2.4	5.76	- 2.4	5.76
26	29	27	- 4.4	19.36	- 1.4	1.96	- 3.4	11.56
28	32	34	- 2.4	5.76	+ 1.6	2.56	+ 3.6	12.96
30	31	33	- 0.4	0.16	+ 0.6	0.36	+ 2.6	6.76
26	32	32	- 4.4	19.36	+ 1.6	2.56	+ 1.6	2.56
29	34	33	- 1.4	1.96	+ 3.6	12.96	+ 2.6	6.76
31	29	29	+ 0.6	0.36	- 1.4	1.96	- 1.4	1.96
32		34	+ 1.6	2.56			+ 3.6	12.96
		33					+ 2.6	6.76
Σ 280	279	352		109.60		41.24		102.16

$$n_1 = 10 ; \quad n_2 = 9 ; \quad n_3 = 11 ; \quad N = n_1 + n_2 + n_3 = 10 + 9 + 11 = 30.$$

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{280}{10} = 28, \quad \bar{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{279}{9} = 31, \quad \bar{X}_3 = \frac{\Sigma X_3}{n_3} = \frac{352}{11} = 32.$$

$$\bar{X} = \frac{\Sigma X_1 + \Sigma X_2 + \Sigma X_3}{N} = \frac{280 + 279 + 352}{30} = 30.4.$$

$$\begin{aligned} SS_t &= \Sigma(X_1 - \bar{X})^2 + \Sigma(X_2 - \bar{X})^2 + \Sigma(X_3 - \bar{X})^2 \\ &= 109.60 + 41.24 + 102.16 = 253.0 ; \quad df_t = N - 1 = 30 - 1 = 29. \end{aligned}$$

$$\begin{aligned} SS_b &= n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 + n_3(\bar{X}_3 - \bar{X})^2 \\ &= 10(28 - 30.4)^2 + 9(31 - 30.4)^2 + 11(32 - 30.4)^2 = 89.0 ; \end{aligned}$$

$$df_b = k - 1 = 3 - 1 = 2.$$

$$SS_w = SS_t - SS_b = 253.0 - 89.0 = 164.0 ; \quad df_w = N - k = 30 - 3 = 27.$$

Alternatively, the abstract reasoning scores are entered in Table 11.12, and used in working out ΣX_1 , ΣX_2 , ΣX_3 , ΣX_1^2 , ΣX_2^2 and ΣX_3^2 .

Table 11.12. Table for computing sums of squares directly from raw abstract reasoning scores.

X_1	X_2	X_3	X_1^2	X_2^2	X_3^2
26	30	34	676	900	1156
27	34	35	729	1156	1225
25	28	28	625	784	784
26	29	27	676	841	729
28	32	34	784	1024	1156
30	31	33	900	961	1089
26	32	32	676	1024	1024
29	34	33	841	1156	1089
31	29	29	961	841	841
32		34	1024		1156
		33			1089
Σ 280	279	352	7892	8687	11338

$$SS_t = \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 - \frac{(\Sigma X_1 + \Sigma X_2 + \Sigma X_3)^2}{N}$$

$$= 7892 + 8687 + 11338 - \frac{(280 + 279 + 352)^2}{30} = 252.97 ;$$

$$df_t = N - 1 = 30 - 1 = 29.$$

$$SS_b = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} - \frac{(\Sigma X_1 + \Sigma X_2 + \Sigma X_3)^2}{N}$$

$$= \frac{(280)^2}{10} + \frac{(279)^2}{9} + \frac{(352)^2}{11} - \frac{(280 + 279 + 352)^2}{30} = 88.97 ;$$

$$df_b = k - 1 = 3 - 1 = 2.$$

$$SS_w = SS_t - SS_b = 252.97 - 88.97 = 164.0 ; \quad df_w = N - k = 30 - 3 = 27.$$

(b) Computation of variances and F ratio :

$$s_b^2 = \frac{SS_b}{df_b} = \frac{88.97}{2} = 44.49 ; \quad s_w^2 = \frac{SS_w}{df_w} = \frac{164.0}{27} = 6.07 ; \quad F = \frac{s_b^2}{s_w^2} = \frac{44.49}{6.07} = 7.33.$$

Table 11.13. Anova table for abstract reasoning scores.

Sources of variation	Sums of squares	df	Variances	F
Between groups	88.97	2	44.49	7.33
Within groups	164.0	27	6.07	
Total	252.97	29		

STATISTICS IN BIOLOGY AND PSYCHOLOGY

c) Significance of F :

$\alpha = 0.01$. Critical $F_{.01(2,27)} = 5.49$ (Table H of Appendix).

As the computed F is higher than the critical F for the 0.01 level of significance, it is significant (< 0.01) — there is a *significant added variance component* (s_a^2) between the groups.

(d) Added variance component :

$$s_a^2 = \frac{s_b^2 - s_w^2}{\frac{1}{k-1} \left[n_1 + n_2 + n_3 - \frac{n_1^2 + n_2^2 + n_3^2}{n_1 + n_2 + n_3} \right]} = \frac{44.49 - 6.07}{\frac{1}{3-1} \left[10 + 9 + 11 - \frac{10^2 + 9^2 + 11^2}{10 + 9 + 11} \right]} = 3.85.$$

Proportionate variation due to s_a^2 is given by :

$$\frac{s_a^2}{s_w^2 + s_a^2} = \frac{3.85}{6.07 + 3.85} = 0.39.$$

So, a proportion of 0.39 of the variance of abstract reasoning scores due to the uncontrolled factors is accounted for by the added variance component.

Example 11.4.3.

The lengths (mm) from the anterior end of the narial opening to the tip of the bony beak in three groups of pigeons, drawn from three different habitats, are given in the first three columns of Table 11.14. Find whether or not the groups are homogeneous.

Solution :

To test for the homogeneity of variances of the groups of pigeons exposed to the uncontrolled independent variable, viz., habitat, a one-way model II anova is applied.

Table 11.14. Table for computing sums of squares from raw scores.

Group 1 (X_1)	Group 2 (X_2)	Group 3 (X_3)	X_1^2	X_2^2	X_3^2
5.3	5.0	4.5	28.09	25.00	20.25
5.9	4.6	4.2	34.81	21.16	17.64
3.9	4.1	3.3	15.21	16.81	10.89
3.8	3.5	3.7	14.44	12.25	13.69
4.7	3.7	3.4	22.09	13.69	11.56
4.4	3.8	3.1	19.36	14.44	9.61
5.7	4.4	4.2	32.49	19.36	17.64
3.6	5.2	3.8	12.96	27.04	14.44
4.5	3.4	4.0	20.25	11.56	16.00
6.1	4.0	5.0	37.21	16.00	25.00
Σ 47.9	41.7	39.2	236.91	177.31	156.72

(a) Partitioning of sums of squares :

The raw scores entered in the first three columns of Table 11.14 are used in working out ΣX_1 , ΣX_2 , ΣX_3 , ΣX_1^2 , ΣX_2^2 and ΣX_3^2 .

$$n_1 = 10 ; \quad n_2 = 10 ; \quad n_3 = 10 ; \quad N = n_1 + n_2 + n_3 = 10 + 10 + 10 = 30.$$

$$SS_t = \Sigma X_1^2 + \Sigma X_2^2 + \Sigma X_3^2 - \frac{(\Sigma X_1 + \Sigma X_2 + \Sigma X_3)^2}{N}$$

$$= 236.91 + 177.31 + 156.72 - \frac{(47.9 + 41.7 + 39.2)^2}{30} = 17.96 ;$$

$$df_t = N - 1 = 30 - 1 = 29.$$

$$SS_b = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \frac{(\Sigma X_3)^2}{n_3} - \frac{(\Sigma X_1 + \Sigma X_2 + \Sigma X_3)^2}{N}$$

$$= \frac{(47.9)^2}{10} + \frac{(41.7)^2}{10} + \frac{(39.2)^2}{10} - \frac{(47.9 + 41.7 + 39.2)^2}{30} = 4.01 ;$$

$$df_b = k - 1 = 3 - 1 = 2.$$

$$SS_w = SS_t - SS_b = 17.96 - 4.01 = 13.95 ; \quad df_w = N - k = 30 - 3 = 27.$$

(b) Computation of variances and F ratio :

$$s_b^2 = \frac{SS_b}{df_b} = \frac{4.01}{2} = 2.01 ; \quad s_w^2 = \frac{SS_w}{df_w} = \frac{13.95}{27} = 0.52. \quad F = \frac{s_b^2}{s_w^2} = \frac{2.01}{0.52} = 3.87.$$

Table 11.15. Anova table for pigeon narial opening data.

Sources of variation	Sums of squares	df	Variances	F
Between groups	4.01	2	2.01	3.87
Within groups	13.95	27	0.52	
Total	17.96	29		

(c) Significance of F :

Critical F values ($df_b = 2$, $df_w = 27$) are quoted from Table H of Appendix.

$$F_{.05(2,27)} = 3.35 ; \quad F_{.01(2,27)} = 5.49.$$

As the computed F exceeds the critical $F_{.05}$, the computed F is considered significant ($P < 0.05$). Thus, there is a significant added variance component between the groups and the groups are not homogeneous. Because all the groups are of equal size ($n = 10$),

$$s_a^2 = \frac{s_b^2 - s_w^2}{n} = \frac{2.01 - 0.52}{10} = 0.15 ; \quad \frac{s_a^2}{s_w^2 + s_a^2} = \frac{0.15}{0.52 + 0.15} = 0.22.$$

Thus, a proportion of 0.22 of the variance of the dependent variable due to uncontrolled factors is accounted for by the added variance component.

11.5 KRUSKAL-WALLIS NONPARAMETRIC ANOVA

This is an efficient rank-dependent non-parametric method of one-way anova for finding the significance of difference between the observations of two or more groups at a time. The Kruskal-Wallis statistic H , when computed from only two groups, is identical with the square of the z score computed from the Mann-Whitney U from the same two groups.

Assumptions

For the Kruskal-Wallis anova, it should be justifiable to assume that :

(a) each score of the dependent variable occurs in the sample at random and independent of all other scores — in other words, the error terms of the scores are independent of each other ;

(b) either the dependent variable is an ordinal one with its magnitudes already given in ranks for the individuals of the sample, or it is a continuous or discrete measurement variable whose scores can be converted into ranks. This assumption makes the test unsuitable for attributes or qualitative variables.

It may be applied to continuous, discrete or ordinal variables, normal or non-normal distributions and small samples, because it does not require the assumptions for continuous and normal distributions of the dependent variable scores.

Computation

(a) Ranks are given in an ascending order to the scores of all the groups taken together. Two or more identical (tied) scores, occurring either in the same group or in different groups, are each given an *average rank* which is the mean of the actual ranks they would have got if they were consecutive non-identical scores. The score, next higher than the tied set, is allotted the rank that it would have occupied if

the tied scores preceding it had separate consecutive ranks instead of an average one. In *Example 11.5.1*, for instance, the score 7.8 occurs once in each of the groups 1 and 2. These two scores would have ranks of 6 and 7 if they were consecutive scores and not tied. So, each of these two tied scores is given an average rank of $(6 + 7)/2$ or 6.5 while the next higher score, viz., 8.2 in group 1, receives the rank of 8 (Table 11.16).

Being a rank statistic, the Kruskal-Wallis statistic H suffers from *inaccuracies* because (i) an average rank is allotted to all the scores of a tied set instead of their true separate ranks, and (ii) successive ranks are given to consecutive scores with no consideration of the varying differences in magnitude between those scores.

(b) The ranks of each group are added separately to give the *rank sums* (R_i) of the respective groups.

(c) The mean rank (\bar{R}_i) of each group and the mean rank (\bar{R}) of all the groups are computed using respectively the size (n_i) of each group and the total size (N) of all the groups ($N = \sum n_i$).

$$\bar{R}_i = \frac{R_i}{n_i}, \quad \bar{R} = \frac{\sum R_i}{\sum n_i} = \frac{\sum R_i}{N}$$

(d) Where there are k number of groups,

$$N = n_1 + n_2 + \dots + n_k ; \quad df = k - 1.$$

$$H = \frac{12 \sum n_i (\bar{R}_i - \bar{R})^2}{N(N+1)} \\ = \frac{12}{N(N+1)} \left[n_1 (\bar{R}_1 - \bar{R})^2 + n_2 (\bar{R}_2 - \bar{R})^2 + \dots \right. \\ \left. \dots + n_k (\bar{R}_k - \bar{R})^2 \right].$$

Alternatively, omitting steps (c) and (d),

$$H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N+1) \\ = \frac{12}{N(N+1)} \left[\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \dots + \frac{R_k^2}{n_k} \right] - 3(N+1).$$

Significance

The null hypothesis (H_0) contends that the computed H is not significantly different from 0, and that there is no significant difference between the scores (or ranks) of the different groups. In other words, the H_0 contends that the groups come from the same population and so, their medians are identical. If the H_0 is correct, the distribution of the statistic H would be almost identical with the chi square distribution for the same df . To find the probability (P) of the H_0 being correct, the computed H is compared with the critical χ^2 for the chosen significance level (α). The computed H is considered *significant* and the groups are considered to *differ significantly*, only if the computed H exceeds or equals the critical χ^2 value for the chosen α ($P \leq \alpha$). If the critical χ^2 is higher than the computed H , there is *no significant difference* between the

scores (or ranks) of the groups ($P > \alpha$).

Multiple comparison by Mann-Whitney test

If the computed H is found significant ($P \leq \alpha$) and there are more than two groups in the experiment, multiple comparison Mann-Whitney U test is performed as an *a-priori* test between groups chosen while designing the experiment, to find which of those groups differ significantly from each other. But when the Kruskal-Wallis test has been done between *only two groups*, a significant H leads directly to the inference that the two groups have a significant difference; no multiple comparison Mann-Whitney test needs to follow.

Where the computed H is not significant ($P > \alpha$), the groups do not differ significantly. In this case also, no multiple comparison test needs to be worked out.

Example 11.5.1.

Three groups of anemic patients were administered three different levels of an anti-anemic factor, after which their blood hemoglobin (g/dL) were found to be as follows.

Group 1 :	7.3,	8.2,	8.4,	7.2,	7.0,	7.6,	7.8,	7.4.	($n_1 = 8$).
Group 2 :	10.4,	7.8,	9.6,	9.2,	10.0,	8.8,	9.4,	9.2.	($n_2 = 8$).
Group 3 :	12.0,	10.6,	13.4,	14.6,	14.0,	14.0,	11.0,	12.0.	($n_3 = 8$).

Apply Kruskal-Wallis test to find whether or not the scores of different groups differ significantly. Also find whether there are significant differences between groups 1 and 2, and between groups 2 and 3.

Solution :

For Kruskal-Wallis nonparametric anova, the hemoglobin scores of the groups are entered in the first, third and fifth columns, respectively, of Table 11.16.

(a) Ranks are assigned in an ascending order to the scores of all three groups taken together. Average ranks are allotted to the scores of each tied set.

(b) The ranks of each group are added separately to give the rank sum (R_i) of that group.

(c) The mean rank (\bar{R}_i) of each group as well as the mean rank (\bar{R}) of all the groups, taken together, is computed.

Table 11.16. Table for nonparametric anova of blood hemoglobin data.

Group 1		Group 2		Group 3	
Scores (X_1)	Ranks	Scores (X_2)	Ranks	Scores (X_3)	Ranks
7.3	3	10.4	16	12.0	19.5
8.2	8	7.8	6.5	10.6	17
8.4	9	9.6	14	13.4	21
7.2	2	9.2	11.5	14.6	24
7.0	1	10.0	15	14.0	22.5
7.6	5	8.8	10	14.0	22.5
7.8	6.5	9.4	13	11.0	18
7.4	4	9.2	11.5	12.0	19.5
Rank sums	38.5 (R_1)		97.5 (R_2)		164.0 (R_3)
Mean ranks	4.81 (\bar{R}_1)		12.19 (\bar{R}_2)		20.5 (\bar{R}_3)

$$\bar{R}_1 = \frac{R_1}{n_1} = \frac{38.5}{8} = 4.81 ; \quad \bar{R}_2 = \frac{R_2}{n_2} = \frac{97.5}{8} = 12.19 ; \quad \bar{R}_3 = \frac{R_3}{n_3} = \frac{164.0}{8} = 20.5 ;$$

$$\bar{R} = \frac{\sum R_i}{\sum n_i} = \frac{38.5 + 97.5 + 164.0}{8 + 8 + 8} = 12.50.$$

(d) The statistic H is computed using the mean ranks.

$$N = n_1 + n_2 + n_3 = 8 + 8 + 8 = 24. \quad df = k - 1 = 3 - 1 = 2.$$

$$H = \frac{12}{N(N+1)} [n_1(\bar{R}_1 - \bar{R})^2 + n_2(\bar{R}_2 - \bar{R})^2 + n_3(\bar{R}_3 - \bar{R})^2]$$

$$= \frac{12}{24 \times 25} [8(4.81 - 12.50)^2 + 8(12.19 - 12.50)^2 + 8(20.5 - 12.50)^2] = 19.72.$$

[Alternatively, H may be computed directly from R_i scores, omitting steps (c) and (d).

$$H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N+1) = \frac{12}{N(N+1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3} \right) - 3(N+1)$$

$$= \frac{12}{24 \times 25} \left(\frac{38.5^2}{8} + \frac{97.5^2}{8} + \frac{164.0^2}{8} \right) - 3 \times 25 = 19.71.]$$

(e) Critical χ^2 values with 2 degrees of freedom are quoted from Table C of Appendix.

$$\chi_{0.05(2)}^2 = 5.99 ; \quad \chi_{0.01(2)}^2 = 9.21 ; \quad \chi_{0.001(2)}^2 = 13.82.$$

As the computed H exceeds the critical χ^2 for 0.001 level, the probability of the null hypothesis being correct is considered too low ($P \ll 0.001$). Thus, there are *significant differences* between the groups.

(f) Multiple comparison Mann-Whitney U test is performed to explore the significance of difference between groups 1 and 2, and between groups 2 and 3. (See § 9.6 for details.)

Groups 1 and 2 :

(i) Scores of these two groups are entered in Table 11.17, ranks are assigned to them in an ascending order taking both the groups together, and the rank sums (R_1 and R_2) are computed for the two groups separately.

(ii) Any of the two rank sums is used for computing the statistic U and for converting it to z . Thus,

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 8 \times 8 + \frac{8 \times 9}{2} - 38.5 = 61.5.$$

$$U_e = \frac{n_1 n_2}{2} = \frac{8 \times 8}{2} = 32.0 ; \quad s_u = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{8 \times 8 (8 + 8 + 1)}{12}} = 9.52 ;$$

$$z = \frac{U_1 - U_e}{s_u} = \frac{61.5 - 32.0}{9.52} = 3.10.$$

(iii) The probability P of the correctness of the H_0 is worked out using the normal curve table (Table A of Appendix).

$$P = 2 [0.5000 - (\text{area from } \mu \text{ to computed } z \text{ of } 3.10)] = 2 (0.5000 - 0.4990) = 0.002.$$

As P is too low, the difference between groups 1 and 2 is considered *significant* ($P = 0.002$).

Groups 2 and 3 :

(i) The scores of these two groups are entered in Table 11.18, ranks are assigned to them taking both the groups together, and the rank sums (R_2 and R_3) are computed for the two groups separately.

(ii) Same as in the previous case :

$$U_2 = n_2 n_3 + \frac{n_2(n_2+1)}{2} - R_2 = 8 \times 8 + \frac{8 \times 9}{2} - 36.0 = 64.0. \quad (\text{Table 11.18})$$

$$U_e = \frac{n_2 n_3}{2} = \frac{8 \times 8}{2} = 32.0 ; \quad s_u = \sqrt{\frac{n_2 n_3 (n_2 + n_3 + 1)}{12}} = \sqrt{\frac{8 \times 8 (8 + 8 + 1)}{12}} = 9.52 ;$$

Table 11.17. Mann-Whitney test for groups 1 and 2.

X_1	Ranks	X_2	Ranks
7.3	3	10.4	16
8.2	8	7.8	6.5
8.4	9	9.6	14
7.2	2	9.2	11.5
7.0	1	10.0	15
7.6	5	8.8	10
7.8	6.5	9.4	13
7.4	4	9.2	11.5
Σ	38.5 (R_1)		97.5 (R_2)

Table 11.18. Mann-Whitney test for groups 2 and 3.

X_2	Ranks	X_3	Ranks
10.4	8	12.0	11.5
7.8	1	10.6	9
9.6	6	13.4	13
9.2	3.5	14.6	16
10.0	7	14.0	14.5
8.8	2	14.0	14.5
9.4	5	11.0	10
9.2	3.5	12.0	11.5
Σ	36.0 (R_2)		100.0 (R_3)

$$z = \frac{U_2 - U_e}{s_u} = \frac{64.0 - 32.0}{9.52} = 3.36.$$

(iii) The normal curve table (Table A) is used for finding P .

$$P = 2 [0.5000 - (\text{area from } \mu \text{ to computed } z \text{ of } 3.36)] = 2 (0.5000 - 0.4996) = 0.0008.$$

As P is too low, the difference between groups 2 and 3 is *considered significant* ($P = 0.0008$).

Example 11.5.2.

Find whether or not there is a significant difference between the strengths of kneejerk reflexes (degrees of arc) of the following groups.

Gr. 1 (athletes) : 31, 30, 22, 30, 26, 28, 19, 36, 37.
 Gr. 2 (nonathletes): 35, 26, 14, 20, 11, 14, 21, 31, 27, 24, 10.

Solution :

The scores (X_1 and X_2) of the two groups are entered in the first and third columns, respectively, of Table 11.19.

Table 11.19. Table for nonparametric anova of kneejerk data. ($n_1 = 9$; $n_2 = 11$)

Athletes		Nonathletes	
Kneejerk strengths (X_1)	Ranks	Kneejerk strengths (X_2)	Ranks
31	16.5	35	18
30	14.5	26	10.5
22	8	14	3.5
30	14.5	20	6
26	10.5	11	2
28	13	14	3.5
19	5	21	7
36	19	31	16.5
37	20	27	12
		24	9
		10	1
Rank sums	121.0 (R_1)		89.0 (R_2)
Mean ranks	13.44 (\bar{R}_1)		8.09 (\bar{R}_2)

(a) Ranks are assigned in an ascending order to the scores of both the groups taken together, allotting average ranks to the scores of each tied set.

(b) The rank sum (R_i) and the mean rank (\bar{R}_i) of each group are computed. The mean rank (\bar{R}) of the scores of all the groups is also computed.

$$\bar{R}_1 = \frac{R_1}{n_1} = \frac{121.0}{9} = 13.44 ; \quad \bar{R}_2 = \frac{R_2}{n_2} = \frac{89.0}{11} = 8.09 ; \quad \bar{R} = \frac{\sum R_i}{\sum n_i} = \frac{121 + 89.0}{9 + 11} = 10.5.$$

(c) The statistic H is computed using the mean ranks.

$$N = n_1 + n_2 = 9 + 11 = 20. \quad df = k - 1 = 2 - 1 = 1.$$

$$H = \frac{12}{N(N+1)} [n_1(\bar{R}_1 - \bar{R})^2 + n_2(\bar{R}_2 - \bar{R})^2]$$

$$= \frac{12}{20 \times 21} [9(13.44 - 10.5)^2 + 11(8.09 - 10.5)^2] = 4.05.$$

[Alternatively, H may be computed directly from the rank sums (R_i).

$$H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N+1) = \frac{12}{N(N+1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} \right) - 3(N+1)$$

$$= \frac{12}{20 \times 21} \left(\frac{121.0^2}{9} + \frac{89.0^2}{11} \right) - 3 \times 21 = 4.05.]$$

(d) Critical χ^2 values ($df = 1$) are quoted from Table C of Appendix.

$$\chi_{0.05(1)}^2 = 3.84 ; \quad \chi_{0.02(1)}^2 = 5.41 ; \quad \chi_{0.01(1)}^2 = 6.64.$$

As the computed H is found higher than the critical χ^2 for the 0.05 level, it is considered significant. So, the scores of the two groups differ significantly ($P < 0.05$).

11.6 TWO-WAY ANOVA

A two-way anova is used to investigate the simultaneous effects of two independent variables or factors on a dependent variable. Here, given combinations of levels of two factors are applied on the individuals of the sample. For two-way anova, a *classification table* is framed with the applied levels of one factor represented along its rows, and those of the other represented along its columns (Tables 11.20 and 11.22).

Two-way anova with replications

A two-way anova with replications is worked out when every combination of the two factors — one level of each — has been applied on more than one individual. The effect of each combination of independent variables is given by the replicated observations in a group of individuals. In the classification table, scores of each such replicated group of observations are included in a particular cell meant for a given combination of the factor levels. The variance of scores within these cells gives the

within-cells or within-groups variance (s_w^2).

Each score of the dependent variable, occurring in any cell of the classification table, is represented here by X_{rci} where the subscripts r , c and i stand respectively for the row and the column to which the cell belongs, and the serial position of the score within the cell. Thus, X_{213} is the 3rd score in the cell belonging to row 2 and column 1. The mean scores of the cells, called the *group means*, are represented by \bar{X}_{rc} . Thus, \bar{X}_{21} is the group mean for the cell belonging to row 2 and column 1. The row means are shown as \bar{X}_r , while the column means are indicated by \bar{X}_c ; thus, \bar{X}_2 and \bar{X}_2 are the means for row 2 and column 2 respectively. The grand mean of all the scores is represented by \bar{X} .

Partitioning of sum of squares :

The *total sum of squares* (SS_t) is the sum of squared deviations of all scores (X_{rci}) from \bar{X} and has $(N-1)$ degrees of freedom where N stands for the total number of scores in the

data. The total variance s_t^2 is obtained by dividing SS_t by its df .

$$SS_t = \sum (X_{rci} - \bar{X})^2 = \sum X_{rci}^2 - \frac{(\sum X_{rci})^2}{N};$$

$$df_t = N - 1; \quad s_t^2 = \frac{SS_t}{df_t} = \frac{SS_t}{N-1}.$$

In this two-way anova, SS_t is partitioned into four independent and additive components, viz., between-rows sum of squares (SS_r), between-columns sum of squares (SS_c), interaction sum of squares (SS_i) and within-cells or within-groups sum of squares (SS_w).

$$SS_t = SS_r + SS_c + SS_i + SS_w$$

The *between-rows sum of squares* (SS_r) is obtained by multiplying the sum of squared deviations of the row means (\bar{X}_r) from the grand mean (\bar{X}), with the product of the number of columns (c) and the number (n) of scores per cell. Because there are r number of rows and one df is lost in using \bar{X} as an estimate of the parametric mean, SS_r has the df of $(r - 1)$.

$$SS_r = nc \left[\sum (\bar{X}_r - \bar{X})^2 \right] = \frac{\sum (\sum X_{rc})^2}{nc} - \frac{(\sum X_{rci})^2}{N};$$

$$df_r = r - 1;$$

where $\sum X_{rc}$ is the sum of the scores of each row, $\sum (\sum X_{rc})^2$ is the sum of the squared row sums for r number of rows, and $\sum X_{rci}$ is the sum of all scores in the sample.

The *between-rows variance* (s_r^2) is obtained by dividing SS_r by its df ; it is the variance due to the added effects of the independent variable represented along the rows of the table.

$$s_r^2 = \frac{SS_r}{df_r} = \frac{SS_r}{r-1}; \quad df_r = r - 1.$$

The *between-columns sum of squares* (SS_c) is obtained by multiplying the sum of squared deviations of the column means (\bar{X}_c) from \bar{X} ,

with the product of the number of rows (r) and the number (n) of scores per cell. Because the columns number c and one df is lost in using \bar{X} , $df_c = c - 1$.

$$SS_c = nr \left[\sum (\bar{X}_c - \bar{X})^2 \right] = \frac{\sum (\sum X_{rc})^2}{nr} - \frac{(\sum X_{rci})^2}{N};$$

$$df_c = c - 1;$$

where $\sum X_{rc}$ is the sum of scores of each column, $\sum (\sum X_{rc})^2$ is the sum of squared column sums for c number of columns, and $\sum X_{rci}$ is the sum of all scores in the sample.

The *between-columns variance* (s_c^2) is obtained by dividing SS_c by its df ; it is the variance owing to the added effects of the second independent variable represented along the columns of the table.

$$s_c^2 = \frac{SS_c}{df_c} = \frac{SS_c}{c-1}; \quad df_c = c - 1.$$

To use in the computation of the interaction SS_i , the *sum of squares between cell means* (SS_{rc}) is computed by multiplying the sum of squared deviations of cell means \bar{X}_{rc} from \bar{X} with the number (n) of scores per cell. As the number of cells or cell means equals rc and one df is lost in using \bar{X} , SS_{rc} has the df of $(rc - 1)$.

$$SS_{rc} = n \left[\sum (\bar{X}_{rc} - \bar{X})^2 \right] = \frac{\sum (\sum X_{rc})^2}{n} - \frac{(\sum X_{rci})^2}{N};$$

$$df_{rc} = rc - 1;$$

where $\sum X_{rc}$ is the sum of scores of each cell and $\sum (\sum X_{rc})^2$ is the sum of squared cell sums.

In higher orders of anova, there is a possibility of *interaction* between the independent variables, either enhancing the effect of one of them on the dependent variable due to the simultaneous effect of another (*synergism*), or reducing the effect of one owing to that of the other (*interference*). *Interaction sum of squares* (SS_i) and *interaction*

variance (s_i^2) are the measures of variations of scores of the dependent variable due to such joint effects of more than one independent variable.

$$SS_i = SS_{rc} - SS_c - SS_r \\ = n[\Sigma(\bar{X}_{rc} - \bar{X}_c - \bar{X}_r + \bar{X})^2];$$

$$df_i = (r-1)(c-1);$$

$$s_i^2 = \frac{SS_i}{df_i} = \frac{SS_i}{(r-1)(c-1)}.$$

The *within-cells* or *residual sum of squares* (SS_w) is that component of SS_i which is left after partitioning off other SS components with known sources. It is obtained as the sum of squared deviations of individual scores (X_{rci}) from the respective cell means (\bar{X}_{rc}), or by subtracting the sum of SS_r , SS_c and SS_i from SS_i . As it involves the total N number of scores and one df is lost in keeping rc number of cell means unchanged, SS_w has the df of $(N-rc)$.

$$SS_w = \Sigma(X_{rci} - \bar{X}_{rc})^2 = SS_i - SS_r - SS_c - SS_i; \\ df_w = N - rc = rc(n-1).$$

The *within-cells variance* (s_w^2) is called the *error variance*, *residual variance* or *remainder variance* as, unlike other variance estimates, its source is not known. It is a measure of the uncontrolled variations of individual scores due to random effects of sampling.

$$s_w^2 = \frac{SS_w}{df_w} = \frac{SS_w}{N-rc}.$$

F ratio in two-way model I anova :

In a model I or *fixed model anova* involving two treatment variables, F ratios are computed to test the significance of s_r^2 , s_c^2 and s_i^2 , using s_w^2 as the error term in the denominator in each case. The df_1 and df_2 of each F ratio are those of its numerator variance and denominator variance, respectively; again, each variance has the same df as the SS used in its computation.

$$F_r = \frac{s_r^2}{s_w^2}; \quad df_1 = df_r = r-1;$$

$$df_2 = df_w = N-rc.$$

$$F_c = \frac{s_c^2}{s_w^2}; \quad df_1 = df_c = c-1;$$

$$df_2 = df_w = N-rc.$$

$$F_i = \frac{s_i^2}{s_w^2}; \quad df_1 = df_i = (r-1)(c-1);$$

$$df_2 = df_w = N-rc.$$

A computed F is considered significant, only if it exceeds or equals the critical F (with the same df_1 and df_2 as the computed F) for the chosen level of significance ($P \leq \alpha$). A significant F_r or F_c denotes a significant change in the dependent variable due to the effect of the treatment variable represented along the rows or the columns, respectively. A significant F_i implies a significant interaction between the two treatment variables.

In case the F ratio of a variance estimate is found significant, *omega square* is computed for that variance estimate to measure what proportion of the variance of dependent variable scores is related to the corresponding effect — row, column or interaction.

$$w_c^2 = \frac{SS_r - (r-1)s_w^2}{s_w^2 + SS_i}; \quad w_r^2 = \frac{SS_c - (c-1)s_w^2}{s_w^2 + SS_i};$$

$$w_i^2 = \frac{SS_i - (r-1)(c-1)s_w^2}{s_w^2 + SS_i}.$$

F ratio in two-way model II anova :

In a model II or *random model anova* involving two classification variables, both beyond the control of the investigator, the F ratio is first computed as F_i for testing the significance of the interaction effect (s_i^2), using s_w^2 as the error term in the denominator. Then, the F ratios for s_r^2 (row effect) and s_c^2 (column effect) are computed with s_i^2 as the denominator variance.

$$F_i = \frac{s_i^2}{s_w^2}; \quad F_r = \frac{s_r^2}{s_i^2}; \quad F_c = \frac{s_c^2}{s_i^2}.$$

As s_w^2 is ordinarily lower than s_i^2 , use of the latter as the error term yields smaller and less significant F ratios and decreases the chances of type I error of inference.

F ratio in two-way model III anova :

In a two-way *mixed model anova* involving one treatment variable and one random classification variable, the denominator variance or error term used in computing the F ratio depends on the following arrangements of the independent variables along rows and columns.

F ratios	rows	columns
$F_r = \frac{s_r^2}{s_w^2}$	random	treatment
$F_r = \frac{s_r^2}{s_i^2}$	treatment	random
$F_c = \frac{s_c^2}{s_w^2}$	treatment	random
$F_c = \frac{s_c^2}{s_i^2}$	random	treatment
$F_i = \frac{s_i^2}{s_w^2}$	any arrangement	

Two-way anova without replication

A two-way anova without replication is used when every combination of the *two factors* — one level of each — has been applied on only one individual. Thus, the effect of each such combination of independent variables is given here by a single observation only.

Each individual score of the dependent variable is represented here by X_{rc} where the subscripts r and c stand respectively for the row and the column to which the score belongs. Thus, X_{23} is the score belonging to

row 2 and column 3 of the classification table. The row means and the column means of the scores are represented respectively by \bar{X}_r and \bar{X}_c ; thus, \bar{X}_2 is the mean score of the second row and \bar{X}_3 is the mean score of the third column. \bar{X} is the grand mean of all the scores.

Partitioning of sum of squares :

The total sum of squares (SS_t) is computed as the sum of squared deviations of all the scores (X_{rc}) from the grand mean (\bar{X}). The total variance (s_t^2) is computed from SS_t .

$$SS_t = \sum (X_{rc} - \bar{X})^2 = \sum X_{rc}^2 - \frac{(\sum X_{rc})^2}{N};$$

$$df_t = N - 1; \quad s_t^2 = \frac{SS_t}{df_t} = \frac{SS_t}{N - 1}.$$

The computed SS_t is partitioned into SS_r , SS_c and SS_i . Each of the latter is then divided by its df to get the corresponding variance estimate.

The *between-rows sum of squares* (SS_r) is computed by multiplying the sum of squared deviations of the row means (\bar{X}_r) from the grand mean (\bar{X}), with the number (c) of columns. Where $\sum X_{rc}$ is the sum of scores of each row, $\sum (\sum X_{rc})^2$ is the sum of squared row sums for r number of rows, and $\sum X_{rc}$ is the sum of all scores,

$$SS_r = c \sum (\bar{X}_r - \bar{X})^2 = \frac{\sum (\sum X_{rc})^2}{c} - \frac{(\sum X_{rc})^2}{N};$$

$$df_r = r - 1.$$

The *between-rows variance* (s_r^2) is a measure of the average variation of row means and is computed from SS_r .

$$s_r^2 = \frac{SS_r}{df_r} = \frac{SS_r}{r - 1}.$$

SS_c is the *between-columns sum of squares*; the *between-columns variance* (s_c^2) measures the average variation of column means. Where

$\sum (\sum X_c)^2$ is the total of the squared sums of column scores for c number of columns,

are computed for testing the significance of s_r^2 and s_c^2 , using s_i^2 as the denominator variance.

$$SS_c = r \sum (\bar{X}_c - \bar{X})^2 = \frac{\sum (\sum X_c)^2}{r} - \frac{(\sum X_{rc})^2}{N};$$

$$F_r = \frac{s_r^2}{s_i^2}; \quad F_c = \frac{s_c^2}{s_i^2}.$$

$$df_c = c - 1; \quad s_c^2 = \frac{SS_c}{df_c} = \frac{SS_c}{c-1}.$$

The residual sum of squares (SS_i) is that component of SS_t which remains after partitioning off SS_r and SS_c . The residual, remainder or interaction variance (s_i^2) includes both the variance due to individual differences associated with random sampling, and the variance owing to interactions, if any, between the independent variables.

$$SS_t = SS_r + SS_c + SS_i.$$

$$\therefore SS_i = SS_t - (SS_r + SS_c);$$

$$df_i = (r-1)(c-1); \quad s_i^2 = \frac{SS_i}{(r-1)(c-1)}.$$

F ratio :

For all three models of such anova, F ratios

Each F ratio is then compared with critical F_α values having df_1 and df_2 identical with those of respectively the numerator variance and the denominator variance used in the computed F .

Both s_r^2 and s_c^2 may be tested in this way for their significance in case of model II anova, irrespective of the significance or otherwise of the interaction effect. But F ratios, computed with s_i^2 as the denominator, cannot be used for testing s_r^2 or s_c^2 in a model I anova unless it can be assumed that s_i^2 is free from interaction effects. In a model III anova, only the variance estimate due to the treatment variable may be tested for significance by computing F with s_i^2 as the denominator, unless it can be assumed that s_i^2 is free from the interaction effect.

Example 11.6.1.

Apply two-way anova without replication to interpret the following data of tracheal ventilations (X ml per minute) of a sample of insects before and after exposure to a hailstorm. ($\alpha = 0.05$.)

Individuals	:	1	2	3	4	5	6	7	8	9	10
Tracheal ventilations :											
(a) before :		75.2	86.0	78.0	75.4	84.5	65.0	72.0	88.0	71.9	68.0
(b) after :		65.2	79.6	80.0	70.4	75.5	74.2	68.0	74.0	59.1	54.0

Solution :

Levels of hailstorm, i.e., before and after exposure to it, constitute a classification or random variable beyond the control of the investigator; on the contrary, individual differences between the insects, resulting from random sampling, may be considered to form a second classification variable. So, a model II two-way anova without replication may be worked out in interpreting the data. For this, the data are arranged in the columns of Table 11.20.

Table 11.20. Classification table for two-way anova of tracheal ventilation data.

Individuals $X_{.c}$	Exposure to hailstorm ($X_{r.}$)		$\Sigma X_{r.}$	$\bar{X}_{r.}$
	Before	After		
1	75.2	65.2	140.4	70.2
2	86.0	79.6	165.6	82.8
3	78.0	80.0	158.0	79.0
4	75.4	70.4	145.8	72.9
5	84.5	75.5	160.0	80.0
6	65.0	74.2	139.2	69.6
7	72.0	68.0	140.0	70.0
8	88.0	74.0	162.0	81.0
9	71.9	59.1	131.0	65.5
10	68.0	54.0	122.0	61.0
$\Sigma X_{.c}$	764.0	700.0	1464.0	73.2
$\bar{X}_{.c}$	76.4	70.0	(ΣX_{rc})	(\bar{X})

(a) The sum of tracheal ventilation scores of each row ($\Sigma X_{r.}$) and that of each column ($\Sigma X_{.c}$) are worked out in Table 11.20 and used in computing the respective row means ($\bar{X}_{r.}$) and column means ($\bar{X}_{.c}$). Thus, the column means, $\bar{X}_{.1}$ and $\bar{X}_{.2}$, are computed using the respective $\Sigma X_{.c}$ values and the numbers (n) of scores of the corresponding columns.

$$\bar{X}_{.1} = \frac{\Sigma X_{.1}}{n} = \frac{764.0}{10} = 76.4; \quad \bar{X}_{.2} = \frac{\Sigma X_{.2}}{n} = \frac{700.0}{10} = 70.0.$$

(b) Similarly, the $\Sigma X_{r.}$ value and the number (n) of scores of each row are used in working out the row mean $\bar{X}_{r.}$ of that row. For example,

$$\bar{X}_{1.} = \frac{\Sigma X_{1.}}{n} = \frac{140.4}{2} = 70.2; \quad \bar{X}_{2.} = \frac{\Sigma X_{2.}}{n} = \frac{165.6}{2} = 82.8.$$

(c) Total sample size (N) is worked out using the numbers of rows (r) and columns (c) while all scores of the sample are totalled to get ΣX_{rc} . These are then used in computing the grand mean (\bar{X}).

$$N = rc = 10 \times 2 = 20. \quad \bar{X} = \frac{\Sigma X_{rc}}{N} = \frac{1464.0}{20} = 73.2.$$

(d) Total sum of squares (SS_t) is worked out and then partitioned into between-rows (SS_r), between-columns (SS_c) and residual or interaction (SS_i) sums of squares. Their respective degrees of freedom are also worked out.

$$SS_t = \Sigma (X_{rc} - \bar{X})^2 = (75.2 - 73.2)^2 + (86.0 - 73.2)^2 + (78.0 - 73.2)^2 + \dots \\ \dots + (74.0 - 73.2)^2 + (59.1 - 73.2)^2 + (54.0 - 73.2)^2 = 1400.32.$$

$$df_t = N - 1 = 20 - 1 = 19.$$

$$SS_r = c \Sigma (\bar{X}_{r.} - \bar{X})^2 = 2 [(70.2 - 73.2)^2 + (82.8 - 73.2)^2 + \dots + (65.5 - 73.2)^2 + (61.0 - 73.2)^2] \\ = 946.66.$$

$$df_r = r - 1 = 10 - 1 = 9.$$

$$SS_c = r \sum (\bar{X}_c - \bar{X})^2 = 10 [(76.4 - 73.2)^2 + (70.0 - 73.2)^2] = 204.80.$$

$$df_c = c - 1 = 2 - 1 = 1.$$

$$SS_i = SS_t - (SS_r + SS_c) = 1400.32 - (946.66 + 204.80) = 248.86.$$

$$df_i = (r - 1)(c - 1) = (10 - 1)(2 - 1) = 9.$$

(e) Between-rows (s_r^2), between-columns (s_c^2) and interaction (s_i^2) variances are next computed using the respective sums of squares and their degrees of freedom.

$$s_r^2 = \frac{SS_r}{df_r} = \frac{946.66}{9} = 105.18; \quad s_c^2 = \frac{SS_c}{df_c} = \frac{204.80}{1} = 204.80; \quad s_i^2 = \frac{SS_i}{df_i} = \frac{248.86}{9} = 27.65.$$

(f) F ratios are next worked out for s_r^2 and s_c^2 .

$$F_r = \frac{s_r^2}{s_i^2} = \frac{105.18}{27.65} = 3.80; \quad df : df_r, df_i = 9, 9.$$

$$F_c = \frac{s_c^2}{s_i^2} = \frac{204.80}{27.65} = 7.41; \quad df : df_c, df_i = 1, 9.$$

Table 11.21. Anova table for tracheal ventilation data.

Sources of variation	Sums of squares	df	Variances	F
Between rows	946.66	9	105.18	3.80
Between columns	204.80	1	204.80	7.41
Interaction	248.86	9	27.65	
Total	1400.32	19		

(g) Critical F scores ($\alpha = 0.05$) are quoted below from Table H and compared respectively with the computed F_r and F_c scores.

$$F_{0.05(9,9)} = 3.18, \text{ for comparing with } F_r;$$

$$F_{0.05(1,9)} = 5.12, \text{ for comparing with } F_c.$$

As the computed F_r is higher than the critical $F_{0.05(9,9)}$, there is a significant effect of hailstorm on tracheal ventilation ($P < 0.05$).

As the computed F_c exceeds the critical $F_{0.05(1,9)}$, there is a significant variance component due to individual variations between the insects of the sample ($P < 0.05$).

Example 11.6.2.

The blood hemoglobin values (g dL^{-1}) were determined in 10 anemic persons before starting any treatment, then after 3 weeks of oral administration of ferrous sulfate, and again after 6 weeks of such administration. The data are presented in the first four columns of Table 11.22. Apply anova to interpret the observed data ($\alpha = 0.01$).

Solution :

The levels (durations) of treatment constitute a treatment variable and the individuals are considered to constitute a random variable. A model III two-way anova without replication may be applied to the data.

Table 11.22. Classification table used for anova of hemoglobin data.

Individuals (X_{rc})	Duration of treatment (X_{rc})			ΣX_{rc}	\bar{X}_{rc}
	0 wk	3 wk	6 wk		
1	8.2	10.4	13.2	31.8	10.60
2	7.2	10.2	14.5	31.9	10.63
3	7.3	9.8	12.4	29.5	9.83
4	8.2	9.3	11.2	28.7	9.57
5	9.4	12.3	14.8	36.5	12.17
6	10.1	12.0	15.2	37.3	12.43
7	11.5	13.4	15.1	40.0	13.33
8	10.2	10.8	12.9	33.9	11.30
9	9.5	10.2	13.3	33.0	11.00
10	8.4	10.6	14.4	33.4	11.13
ΣX_{rc}	90.0	109.0	137.0	336.0	11.20
\bar{X}_{rc}	9.00	10.90	13.70	(ΣX_{rc})	(\bar{X})

(a) The sum of scores of each row (ΣX_{rc}) and that of each column (ΣX_{rc}) are computed in Table 11.22 and used in computing the row mean (\bar{X}_{rc}), the column mean (\bar{X}_{rc}) and the grand mean (\bar{X}). For each row, $n = 3$ and $\bar{X}_{rc} = \Sigma X_{rc}/n$; for each column, $n = 10$ and $\bar{X}_{rc} = \Sigma X_{rc}/n$; for the grand mean, $\bar{X} = \Sigma X_{rc}/N$, each score being represented as X_{rc} . For example,

$$N = rc = 10 \times 3 = 30. \quad \bar{X} = \frac{\Sigma X_{rc}}{N} = \frac{336.0}{30} = 11.20.$$

$$\bar{X}_{.1} = \frac{\Sigma X_{.1}}{n} = \frac{90.0}{10} = 9.00. \quad \bar{X}_{1.} = \frac{\Sigma X_{1.}}{n} = \frac{31.8}{3} = 10.60.$$

$$(b) SS_r = \Sigma (X_{rc} - \bar{X})^2 = (8.2 - 11.2)^2 + (7.2 - 11.2)^2 + \dots + (13.3 - 11.2)^2 + (14.4 - 11.2)^2 = 158.94.$$

$$df_r = N - 1 = 30 - 1 = 29.$$

$$(c) SS_r = c \Sigma (\bar{X}_{rc} - \bar{X})^2 = 3 [(10.6 - 11.2)^2 + \dots + (11.13 - 11.2)^2] = 36.79. \quad df_r = r - 1 = 10 - 1 = 9.$$

$$s_r^2 = \frac{SS_r}{r-1} = \frac{36.79}{9} = 4.09.$$

$$(d) SS_c = r \Sigma (\bar{X}_{rc} - \bar{X})^2 = 10 [(9 - 11.2)^2 + \dots + (13.7 - 11.2)^2] = 111.80. \quad df_c = c - 1 = 3 - 1 = 2.$$

$$s_c^2 = \frac{SS_c}{c-1} = \frac{111.80}{2} = 55.90.$$

$$(e) SS_i = SS_r - (SS_r + SS_c) = 158.94 - (36.79 + 111.80) = 10.35.$$

$$df_i = (r - 1)(c - 1) = (10 - 1)(3 - 1) = 18;$$

$$s_i^2 = \frac{SS_i}{(r-1)(c-1)} = \frac{10.35}{18} = 0.58.$$

(f) Provided it can be assumed that there is no interaction between the two independent variables,

$$F_r = \frac{s_r^2}{s_i^2} = \frac{4.09}{0.58} = 7.05; \quad df: 9, 18. \quad F_c = \frac{s_c^2}{s_i^2} = \frac{55.90}{0.58} = 96.38; \quad df: 2, 18.$$

Table 11.23. Anova table for hemoglobin data.

Sources of variation	Sums of squares	df	Variances	F
Between rows	36.79	9	4.09	7.05
Between columns	111.80	2	55.90	96.38
Remainder	10.35	18	0.58	
Total	158.94	29		

The computed F_r is greater than the critical $F_{.01(9,18)}$ which amounts to 3.60 (Table H). Similarly, the computed F_c is greater than the critical $F_{.01(2,18)}$ which amounts to 6.01 (Table H). Thus, both F_r and F_c are significant ($P < 0.01$). Hence, there is a significant effect of the treatment variable (column effect) as well as a significant added variance component due to the random effects between individuals (row effect).

Example 11.6.3.

Different combinations of three levels of a fixed treatment variable (A) and two levels of another treatment variable (B) were administered to the groups of a sample of 30 individuals to study their effects on a particular dependent variable. The scores of the dependent variable, measured after such treatments, are arranged in a two-way classification table (Table 11.24). Find the significance of the effects of the treatments and also of their interaction effects ($\alpha = 0.01$).

Solution :

A two-way model I anova with replications may be used. Three levels of variable A are represented along the columns c_1 , c_2 , and c_3 (column effect) while two levels of variable B are represented along the rows r_1 and r_2 (row effect) in Table 11.24. Each combination was replicated with 5 individuals ($n = 5$).

Table 11.24. Two-way classification table for scores of the dependent variable at different levels of two "fixed" treatment variables.

Variable B	Variable A		
	c_1	c_2	c_3
r_1	7	10	14
	9	11	17
	8	12	18
	6	9	12
	10	13	19
r_2	9	18	26
	12	17	25
	10	21	32
	8	16	26
	11	18	26
$n = 5.$		$N = nrc = 5 \times 2 \times 3 = 30.$	

The data are repeated in Table 11.25 for computations.

(a) Means are first computed for cells, rows and columns.

1. Cell means (\bar{X}_{rc}) : $\bar{X}_{rc} = \frac{\sum X_{rc}}{n}$.

$$\bar{X}_{11} = \frac{40}{5} = 8 ; \quad \bar{X}_{12} = \frac{55}{5} = 11 ; \quad \bar{X}_{13} = \frac{80}{5} = 16 ;$$

$$\bar{X}_{21} = \frac{50}{5} = 10 ; \quad \bar{X}_{22} = \frac{90}{5} = 18 ; \quad \bar{X}_{23} = \frac{135}{5} = 27.$$

2. Row means (\bar{X}_r) : $\bar{X}_r = \frac{\sum X_r}{3n}$.

$$\bar{X}_1 = \frac{175}{15} = 11.67 ; \quad \bar{X}_2 = \frac{275}{15} = 18.33.$$

3. Column means (\bar{X}_c) : $\bar{X}_c = \frac{\sum X_c}{2n}$.

$$\bar{X}_1 = \frac{90}{10} = 9.0 ; \quad \bar{X}_2 = \frac{145}{10} = 14.5 ; \quad \bar{X}_3 = \frac{215}{10} = 21.5.$$

4. Grand mean (\bar{X}) : $\bar{X} = \frac{\sum X_{rci}}{N}$.

$$\bar{X} = \frac{450}{30} = 15.0.$$

(b) The sums of squares are next computed.

1. $SS_r = nc [\sum (\bar{X}_r - \bar{X})^2] = 5 \times 3 [(11.67 - 15.00)^2 + (18.33 - 15.00)^2] = 332.67.$

$$df_r = r - 1 = 2 - 1 = 1$$

2. $SS_c = nr [\sum (\bar{X}_c - \bar{X})^2] = 5 \times 2 [(9 - 15)^2 + (14.5 - 15)^2 + (21.5 - 15)^2] = 785.00.$

$$df_c = c - 1 = 3 - 1 = 2.$$

3. $SS_{rc} = n [\sum (\bar{X}_{rc} - \bar{X})^2] = 5 [(8 - 15)^2 + (11 - 15)^2 + (16 - 15)^2 + (10 - 15)^2 + (18 - 15)^2 + (27 - 15)^2] = 1220.00.$

$$df_{rc} = rc - 1 = 2 \times 3 - 1 = 5.$$

4. $SS_i = SS_{rc} - SS_c - SS_r = 1220.00 - 785.00 - 332.67 = 102.33.$

$$df_i = (r - 1)(c - 1) = (2 - 1)(3 - 1) = 2.$$

5. $SS_w = \sum (X_{rci} - \bar{X}_{rc})^2 = (7 - 8)^2 + (9 - 8)^2 + (8 - 8)^2 + \dots + (26 - 27)^2 + (26 - 27)^2 = 110.00.$

$$df_w = N - rc = 30 - 2 \times 3 = 24.$$

(c) The variance estimates are computed dividing each SS with its df.

$$s_r^2 = \frac{SS_r}{df_r} = \frac{332.67}{1} = 332.67 ; \quad s_c^2 = \frac{SS_c}{df_c} = \frac{785.00}{2} = 392.50 ;$$

$$s_i^2 = \frac{SS_i}{df_i} = \frac{102.33}{2} = 51.17 ; \quad s_w^2 = \frac{SS_w}{df_w} = \frac{110.00}{24} = 4.58.$$

Table 11.25. Table for computing two-way model I anova.

Variable B	Variable A			Row sums $\Sigma X_{r.}$	$\bar{X}_{r.}$
	c_1	c_2	c_3		
r_1	7	10	14		
	9	11	17		
	8	12	18		
	6	9	12		
	10	13	19		
ΣX_{rc}	40	55	80	175	11.67
\bar{X}_{rc}	8	11	16		
r_2	9	18	26		
	12	17	25		
	10	21	32		
	8	16	26		
	11	18	26		
ΣX_{rc}	50	90	135	275	18.33
\bar{X}_{rc}	10	18	27		
Column sums $\Sigma X_{.c}$	90	145	215	450 (ΣX_{rci})	
$\bar{X}_{.c}$	9	14.5	21.5	15.0 (\bar{X})	

(d) The F ratios are computed for testing the row effect, the column effect and the interaction effect.

$$F_r = \frac{s_r^2}{s_w^2} = \frac{332.67}{4.58} = 72.635 ; \quad F_c = \frac{s_c^2}{s_w^2} = \frac{392.50}{4.58} = 85.699 ; \quad F_i = \frac{s_i^2}{s_w^2} = \frac{51.17}{4.58} = 11.172.$$

(e) Each computed F ratio has two degrees of freedom, viz., those of its numerator and denominator variances. Thus,

$$\text{for } F_r : df_r, df_w = 1, 24 ; \quad \text{for } F_c : df_c, df_w = 2, 24 ; \quad \text{for } F_i : df_i, df_w = 2, 24.$$

So, F_r should be compared with critical $F_{.01(1,24)}$ while F_c and F_i should be compared with critical $F_{.01(2,24)}$ from Table H.

$$\text{Critical } F_{.01(1,24)} = 7.82 ; \quad F_r = 72.64 ; \quad \therefore P < 0.01.$$

$$\text{Critical } F_{.01(2,24)} = 5.61 ; \quad F_c = 85.70 ; \quad \therefore P < 0.01 ;$$

$$F_i = 11.17 ; \quad \therefore P < 0.01.$$

Hence, all the computed F ratios are significant ($P < 0.01$). So, the effects of both treatment variables (row effect and column effect) as well as the effect of their interaction are significant on the dependent variable.

Table 11.26. Anova table for the data of Table 11.24.

Sources of variation	Sums of squares	df	Variances	F
Between rows	332.67	1	332.67	72.635
Between columns	785.00	2	392.50	85.699
Interaction	102.33	2	51.17	11.172
Within cells	110.00	24	4.58	
Total	1330.00 (SS_T)	29		

*(f) The omega squares are computed for all three significant effects, using the total sum of squares (SS_t) worked out in Table 11.26.

$$w_r^2 = \frac{SS_c - (c-1)s_w^2}{s_w^2 + SS_t} = \frac{785 - 2 \times 4.58}{4.58 + 1330} = 0.581 ;$$

$$w_c^2 = \frac{SS_r - (r-1)s_w^2}{s_w^2 + SS_t} = \frac{332.67 - 4.58}{4.58 + 1330} = 0.246 ;$$

$$w_i^2 = \frac{SS_i - (r-1)(c-1)s_w^2}{s_w^2 + SS_t} = \frac{102.33 - 2 \times 4.58}{4.58 + 1330} = 0.070.$$

So, the row effect, the column effect and the interaction effect are related to 0.581, 0.246 and 0.070 proportions, respectively, of the total variance of the dependent variable scores.

GLOSSARY

- added treatment component** : that component of the variance of dependent variable scores in an experiment which occurs between the groups of subjects or cases in addition to the random variations measured by the within-groups variance, and results from the effects of the controlled treatment variable(s) used as the independent variable(s).
- added variance component** : that component of the variance of dependent variable scores in an experiment which occurs between the groups of subjects or cases in addition to the random variations measured by the within-groups variance, and results from the effects of the uncontrolled classification variable(s) used as the independent variable(s).
- after-design/a-posteriori comparison** : multiple comparison test of group means chosen only after scrutiny of the data already collected in the experiment.
- anova** : analysis of variances of the dependent variable scores in an experiment for estimating the relative magnitudes of a variance under investigation and the error variance, for drawing inferences.
- anova, fixed model or model I** : anova to be worked out with the data of an experiment using independent variable(s) of only the "fixed" or controlled treatment class.
- anova, mixed model or model III** : anova to be applied to the data of an experiment using both controlled ("fixed") treatment(s) and uncontrolled classification variable(s) as independent variables.
- anova, one-way** : anova to be worked out with the data of a single-factor experiment with a single independent variable.
- anova, random model or model II** : anova to be worked out with the data of an experiment using independent variable(s) of only the uncontrolled classification type.
- anova, two-way** : anova to be used in case of a factorial experiment based on the application of combinations of two independent variables.
- before-design/a-priori comparison** : pre-planned multiple comparison test of such group means as have already been chosen while designing the experiment and prior to the collection of data.
- controlled experiment** : experiment using "fixed" treatment variable(s) as independent variable(s) which are under the strict control of the investigator and not liable to random errors.

error variance : variance of dependent variable scores, resulting from unknown and uncontrolled factors associated with random sampling, and used as the denominator of the F ratio.

experimental design : scientific planning of an experiment by determining the required sample size, using random sampling, fixing the levels of application of the independent variable and the number of replications of each such level, applying those levels in a random sequence, minimizing experimental errors from unwanted relevant variables, and allowing the subsequent application of anova on the collected data for interpretation.

F ratio : variance ratio in which the numerator is the variance estimate whose source and significance are under investigation, and the denominator is the error variance resulting from unknown factors associated with random sampling.

factorial experiment : experiment designed to study the changes of the dependent variable on exposure to chosen combinations of different levels of more than one independent variable.

homoscedasticity : the assumption that the groups, drawn for an experiment, initially have homogeneous variances which differ only due to sampling errors.

Kruskal-Wallis H : nonparametric statistic computed in the Kruskal-Wallis rank-dependent one-way anova and interpreted using critical chi square values.

level : chosen amounts, intensities, amplitudes, qualities or categories of an independent variable, to which the dependent variable is exposed in an experiment.

multiple comparison test : test to be applied to different pairs of group means, following a significant F ratio in an anova with more than two groups, to find which group means differ significantly from each other.

omega square : statistic, computed if a model I anova has yielded a significant F ratio, to estimate the strength of association between the independent variable and the dependent variable.

prospective method : an investigation in which the application of the independent variable is subsequently followed by the study of the changes in the dependent variable.

replication : that number of individuals or cases of a sample, on which each level of the independent variable is applied to minimize experimental errors.

retrospective method : an investigation to explore the past exposure of the subjects, showing at present specific changes of the dependent variable, to a chosen independent variable.

single-factor experiment : experiment designed to study the changes of the dependent variable due to the exposure of the sample to different levels of a single independent variable.

sum of squares, between-groups : the sum of squares computed by giving the weight of the respective group sizes to the squared deviation of each group mean from the grand mean of a sample.

sum of squares, total : the sum of the squared deviations of the raw scores of all groups in a sample from its grand mean.

sum of squares, within-groups : the sum of the squared deviations of the raw scores of all the groups in a sample, from the respective group means.

uncontrolled experiment : experiment using such independent variable(s) as are beyond the control of the investigator and are consequently liable to random errors.

variance, between-columns : variance due to the added effects of an independent variable represented along the columns of a classification table in a two-way anova.

variance, between-groups : variance of scores of all the groups of subjects in an experiment, computed from the weighted deviations of the respective group means from the grand mean of all the groups.

variance, between-rows : variance owing to the added effects of an independent variable represented along the rows of a classification table in a two-way anova.

variance, interaction : variance due to the joint effects of more than one independent variable in a higher order of anova.

variance, total : variance of scores of all the groups of subjects about their grand mean in an experiment.

variance, within-cells : variance of scores within the cells of a classification table in a two-way anova, serving as a measure of uncontrolled variations of dependent variable scores due to random sampling.

variance, within-groups : variance of scores of all the groups of subjects about the respective group means in an experiment, resulting from unknown and uncontrolled factors due to random sampling.

STEPS IN STATISTICAL TESTS

A significance test

1.

Working out the SE of computed statistic

$$\text{For } r, s_r = \sqrt{\frac{1-r^2}{n-2}}; \text{ for } \bar{X}_1 - \bar{X}_2, s_{\bar{X}_1 - \bar{X}_2} = \sqrt{s_{\bar{X}_1}^2 + s_{\bar{X}_2}^2}.$$

2.

Transforming the computed statistic to a 'standard' score by division with its SE

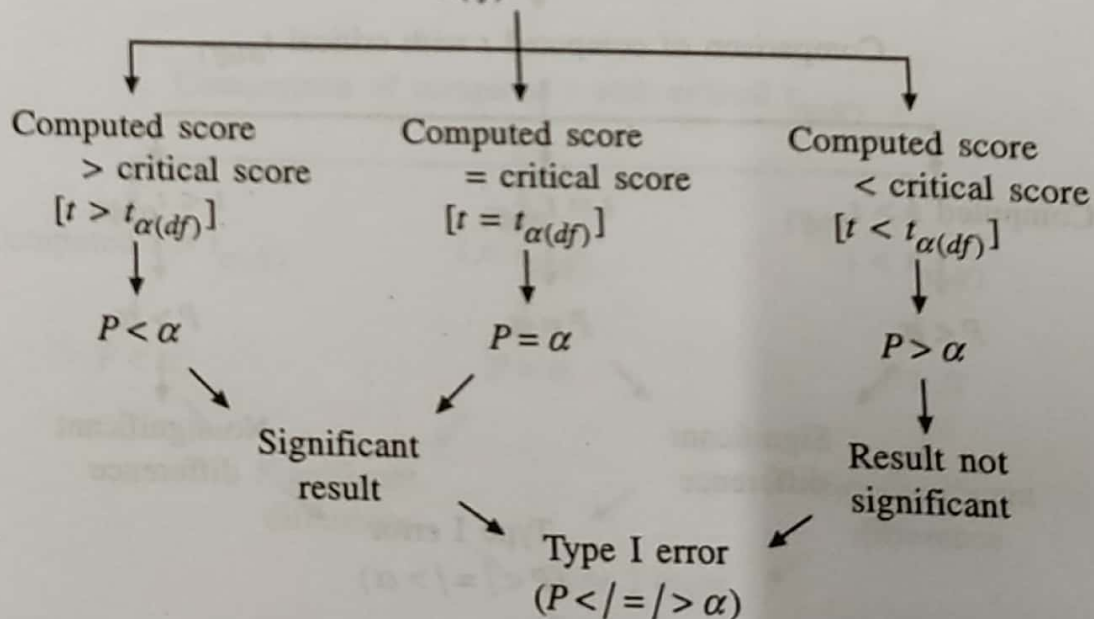
$$\text{E.g., } t = \frac{r}{s_r}; t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}; F = \frac{(\bar{X}_1 - \bar{X}_2)^2}{s_{\bar{X}_1 - \bar{X}_2}^2}.$$

3.

Working out the degrees of freedom

4.

Comparison of the computed 'standard' score (e.g., t and F) with the critical 'standard' score, e.g., $t_{\alpha(df)}$, for the chosen α



5.

t test for $(\bar{X}_1 - \bar{X}_2)$ of large independent groups

1. Construction of Table following formulae



2. Working out of means and sums of squares (SS)

$$\bar{X}_1 = \frac{\sum X_1}{n_1}; \quad \bar{X}_2 = \frac{\sum X_2}{n_2}; \quad SS_1 = \sum (X_1 - \bar{X}_1)^2; \quad SS_2 = \sum (X_2 - \bar{X}_2)^2.$$



3. Computation of SDs and SEs of means

$$s_1 = \sqrt{\frac{\sum (X_1 - \bar{X}_1)^2}{n_1 - 1}}; \quad s_2 = \sqrt{\frac{\sum (X_2 - \bar{X}_2)^2}{n_2 - 1}}; \quad s_{\bar{X}_1} = \frac{s_1}{\sqrt{n_1}}; \quad s_{\bar{X}_2} = \frac{s_2}{\sqrt{n_2}}.$$



4. Working out $s_{\bar{X}_1 - \bar{X}_2}$ and transforming $(\bar{X}_1 - \bar{X}_2)$ into *t*

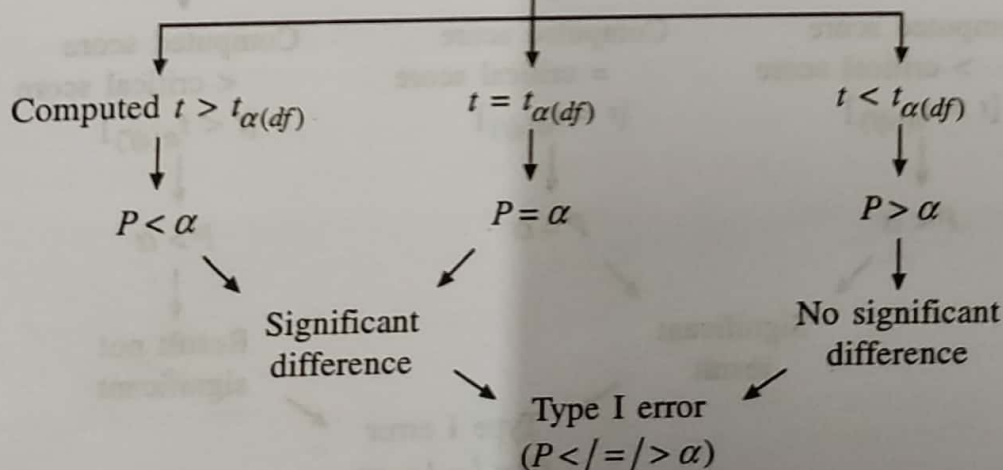
$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{s_{\bar{X}_1}^2 + s_{\bar{X}_2}^2}; \quad t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}; \quad df = n_1 + n_2 - 2.$$



5. Quoting critical $t_{\alpha(df)}$ for chosen α

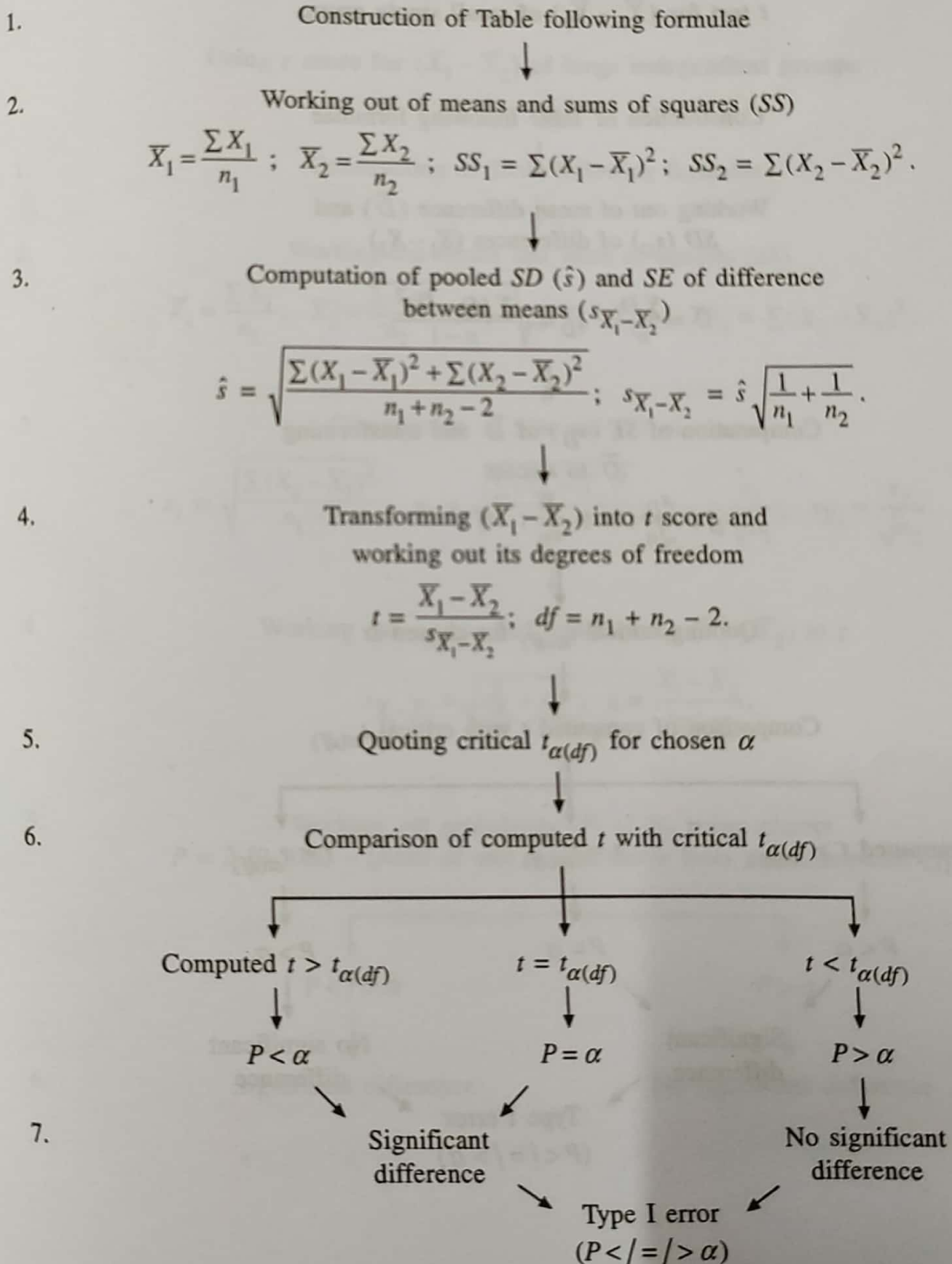


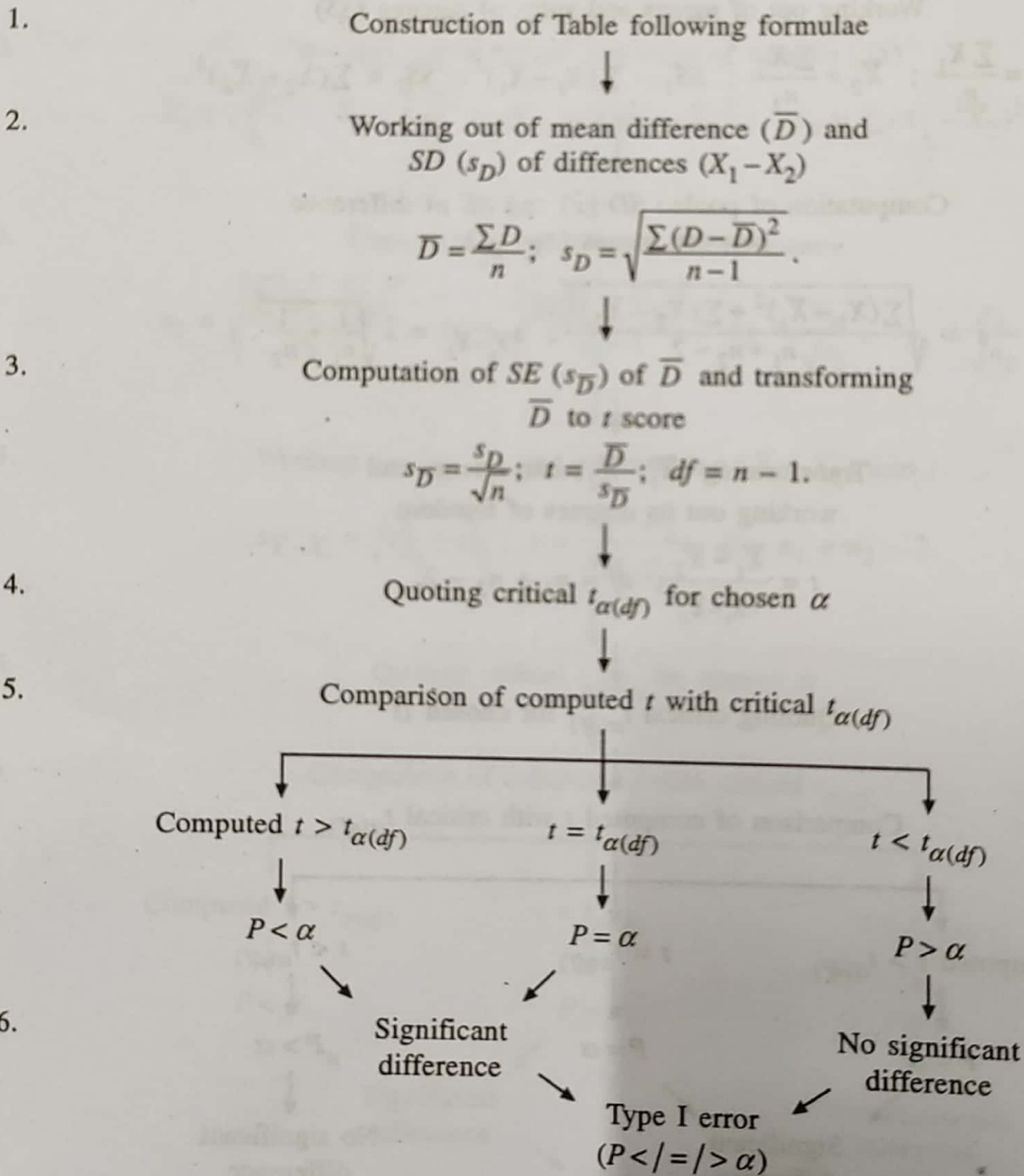
6. Comparison of computed *t* with critical $t_{\alpha(df)}$



- 7.

t test for $(\bar{X}_1 - \bar{X}_2)$ of small independent groups



***t* test for $(\bar{X}_1 - \bar{X}_2)$ of small single group**

Using z score for $(\bar{X}_1 - \bar{X}_2)$ of large independent groups

1. Construction of Table following formulae
↓
2. Working out means and sums of squares (SS)
$$\bar{X}_1 = \frac{\sum X_1}{n_1}; \quad \bar{X}_2 = \frac{\sum X_2}{n_2}; \quad SS_1 = \sum (X_1 - \bar{X}_1)^2; \quad SS_2 = \sum (X_2 - \bar{X}_2)^2.$$

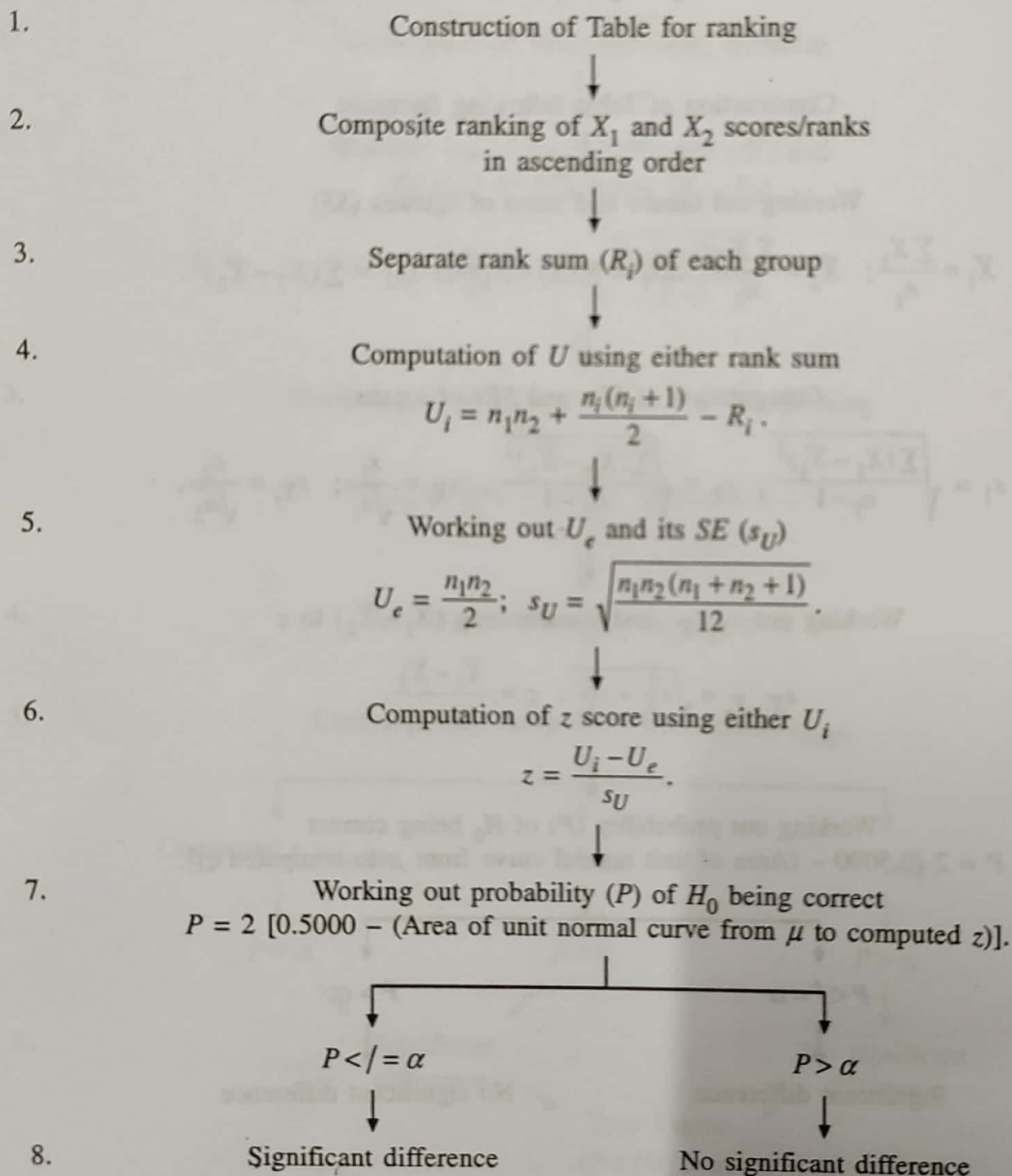
↓
3. Computation of SDs and SEs of means
$$s_1 = \sqrt{\frac{\sum (X_1 - \bar{X}_1)^2}{n_1 - 1}}; \quad s_2 = \sqrt{\frac{\sum (X_2 - \bar{X}_2)^2}{n_2 - 1}}; \quad s_{\bar{X}_1} = \frac{s_1}{\sqrt{n_1}}; \quad s_{\bar{X}_2} = \frac{s_2}{\sqrt{n_2}}.$$

↓
4. Working out $s_{\bar{X}_1 - \bar{X}_2}$ and transforming $(\bar{X}_1 - \bar{X}_2)$ to z
$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{s_{\bar{X}_1}^2 + s_{\bar{X}_2}^2}; \quad z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}.$$

↓
5. Working out probability (P) of H_0 being correct
 $P = 2 [0.5000 - (\text{Area of unit normal curve from } \mu \text{ to computed } z)].$
↓
6.

$P < / = \alpha$
↓
Significant difference

$P > \alpha$
↓
No significant difference

Mann-Whitney U test for $(\bar{X}_1 - \bar{X}_2)$ 

Chi square test for goodness of fit

1. Working out f_e values for k number of classes as per proposed distribution
↓
2. Construction of Table following formulae
↓
3. Entry of f_o and f_e values in respective columns of Table
↓
4. Working out $(f_o - f_e)$, $(f_o - f_e)^2$ and $(f_o - f_e)^2 / f_e$ values
↓
5. Computation of χ^2 and its degrees of freedom

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}; \quad df = k - m,$$

where $m = 1, 2$ or 3 for respectively Mendelian, binomial and normal distributions.

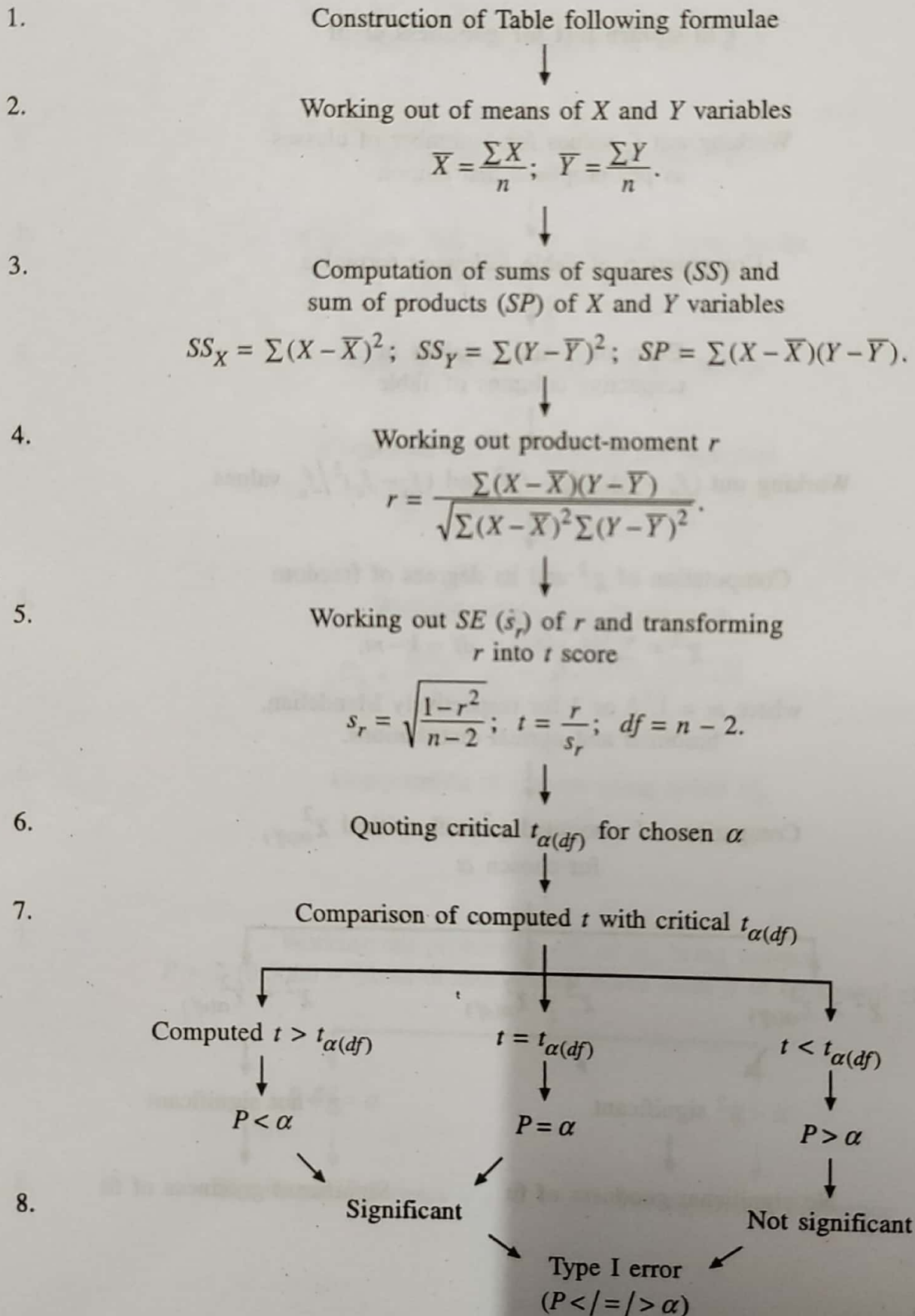
↓
6. Comparison of computed χ^2 with critical $\chi^2_{\alpha(df)}$ for chosen α

$\chi^2 > \chi^2_{\alpha(df)}$
↓
 χ^2 significant
↓
No significant goodness of fit

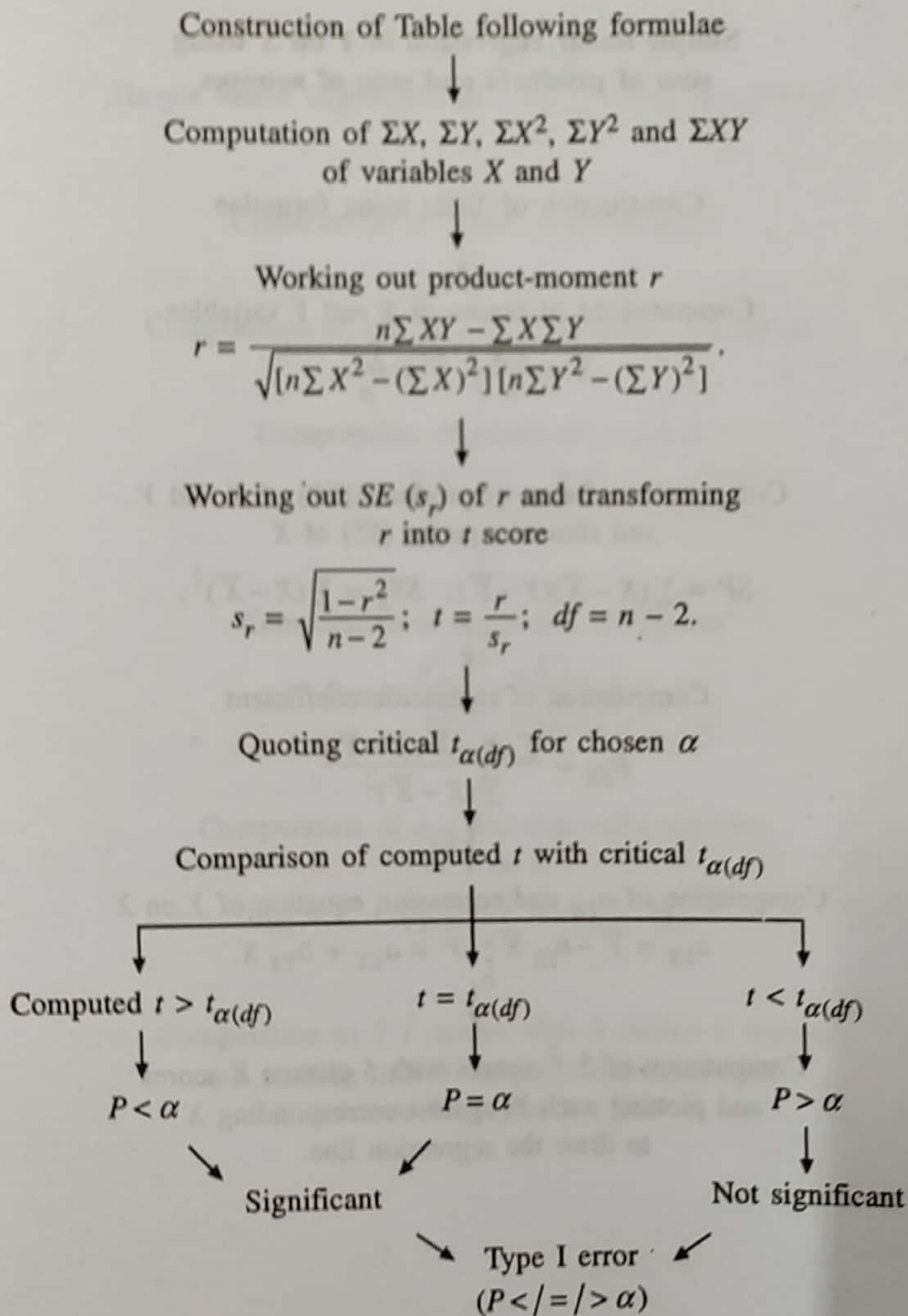
$\chi^2 = \chi^2_{\alpha(df)}$
↓
 χ^2 not significant
↓
Significant goodness of fit

$\chi^2 < \chi^2_{\alpha(df)}$
↓
 χ^2 not significant
↓
Significant goodness of fit
- 7.
- 8.

Product-moment correlation using sum of products



Product-moment correlation using raw scores



**Simple linear regression of Y on X using
sum of products and sum of squares**

1. Construction of Table using formulae
↓
2. Computations of means of X and Y variables

$$\bar{X} = \frac{\sum X}{n}; \quad \bar{Y} = \frac{\sum Y}{n}.$$
↓
3. Computation of sum of products (SP) of X and Y
and sum of squares (SS) of X

$$SP = \sum (X - \bar{X})(Y - \bar{Y}); \quad SS_X = \sum (X - \bar{X})^2.$$
↓
4. Computation of regression coefficient

$$b_{YX} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}.$$
↓
5. Computation of a_{YX} and regression equation of Y on X

$$a_{YX} = \bar{Y} - b_{YX} \bar{X}; \quad \hat{Y} = a_{YX} + b_{YX} X.$$
↓
6. Computation of 5 \hat{Y} scores with 5 chosen X scores
and plotting each \hat{Y} against corresponding X
to draw the regression line.

Simple linear regression of Y on X using raw scores.

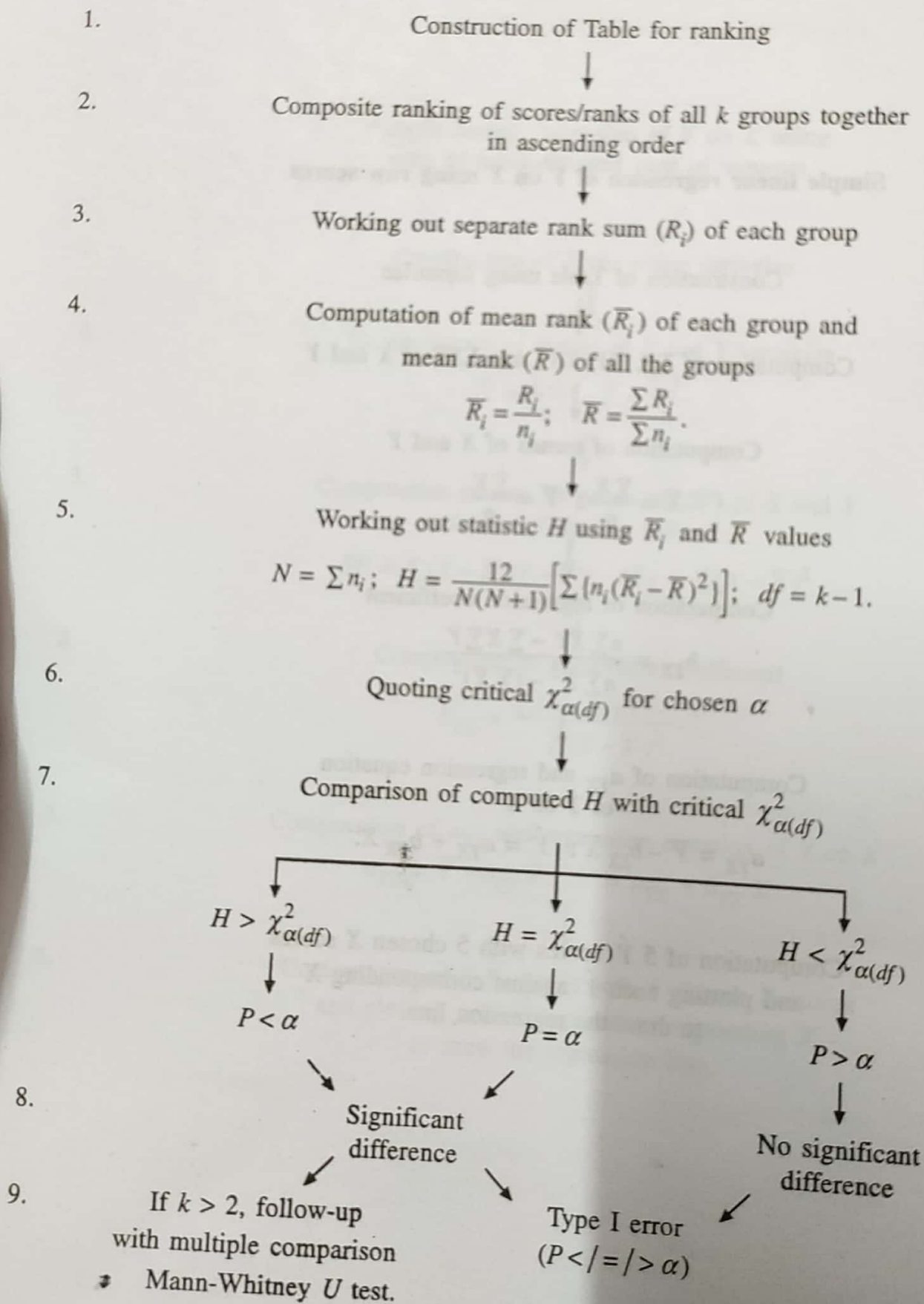
1. Construction of Table using formulae
↓
2. Computation of ΣX , ΣY , ΣX^2 and ΣXY of X and Y
↓
3. Computation of means of X and Y

$$\bar{X} = \frac{\Sigma X}{n}; \quad \bar{Y} = \frac{\Sigma Y}{n}.$$
 ↓
4. Computation of regression coefficient

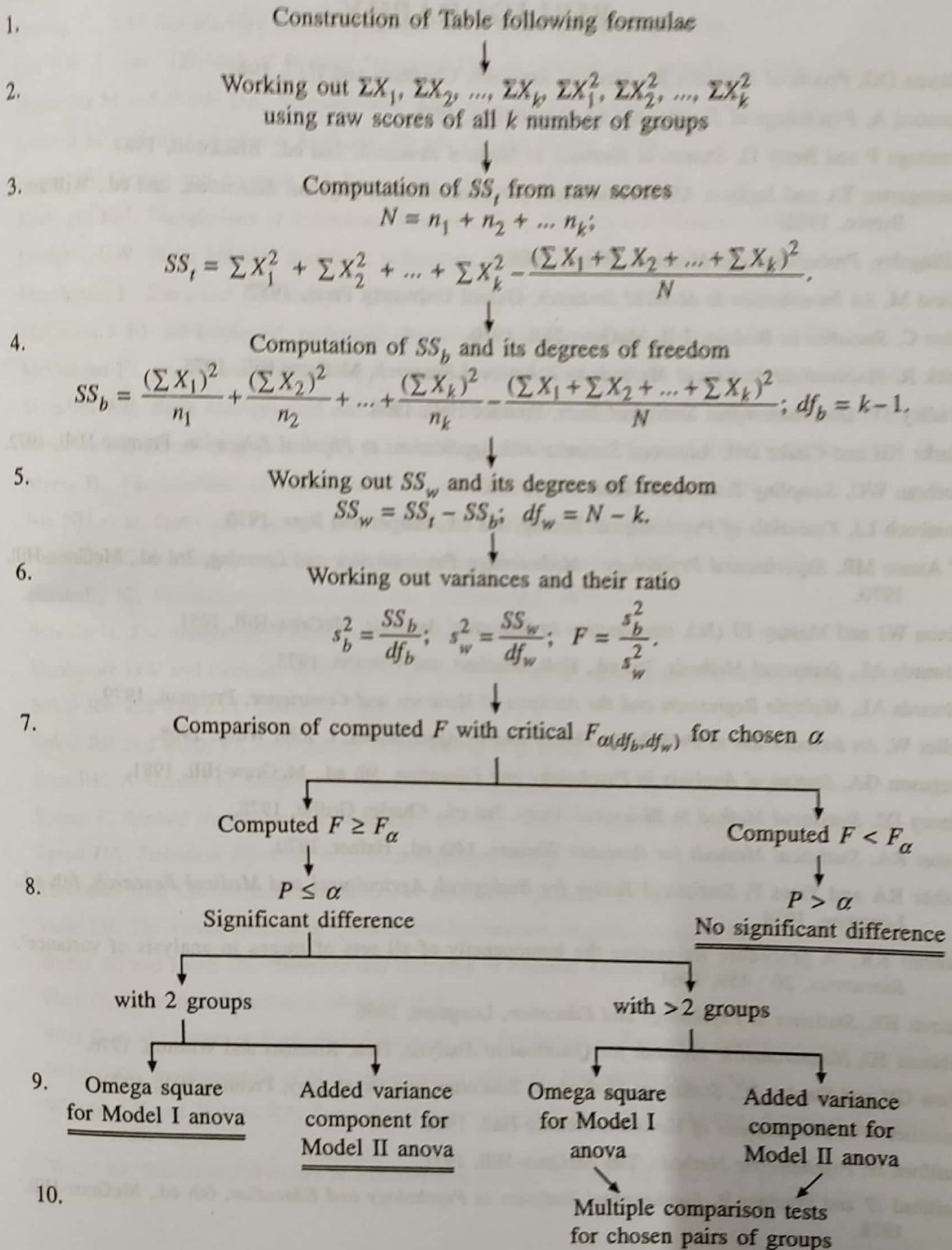
$$b_{YX} = \frac{n \Sigma XY - \Sigma X \Sigma Y}{n \Sigma X^2 - (\Sigma X)^2}.$$
 ↓
5. Computation of a_{YX} and regression equation of Y on X

$$a_{YX} = \bar{Y} - b_{YX} \bar{X}; \quad \hat{Y} = a_{YX} + b_{YX} X.$$
 ↓
6. Computation of 5 \hat{Y} scores with 5 chosen X scores and plotting each \hat{Y} against corresponding X to draw the regression line.

Kruskal-Wallis nonparametric anova



One-way anova



BIBLIOGRAPHY

- Altman DG, *Practical Statistics for Medical Research*, Chapman and Hall, 1991.
- Anastasi A, *Psychological Testing*, 6th ed., Macmillan, 1988.
- Armitage P and Berry G, *Statistical Methods in Medical Research*, 2nd ed., Blackwell, 1987.
- Baumgarten TA and Jackson AS, *Measurement for Evaluation in Physical Education*, 2nd ed., William C. Brown, 1982.
- Billingsley, *Probability and Measure*, Wiley, 1986.
- Bland M, *An Introduction to Medical Statistics*, Oxford University Press, 1987.
- Bliss C, *Statistics in Biology*, I-II, McGraw-Hill, 1970.
- Bock R, *Multivariate Statistical Methods in Behavioral Research*, McGraw-Hill, 1975.
- Bradley JV, *Distribution-free Statistical Tests*, Prentice-Hall, 1968.
- Clarke HH and Clarke DH, *Advanced Statistics with Applications to Physical Education*, Prentice-Hall, 1972.
- Cochran WG, *Sampling Techniques*, John Wiley, 1963.
- Cronbach LJ, *Essentials of Psychological Testing*, 3rd ed., Harper and Row, 1970.
- D' Amato MR, *Experimental Psychology : Methodology, Psychophysics and Learning*, 3rd ed., McGraw-Hill, 1970.
- Dixon WJ and Massey FJ (Jr.), *Introduction to Statistical Analysis*, McGraw-Hill, 1951.
- Edwards AL, *Statistical Methods*, 3rd ed., Holt, Rinehart and Winston, 1973.
- Edwards AL, *Multiple Regression and the Analysis of Variance and Covariance*, Freeman, 1979.
- Feller W, *An Introduction to Probability Theory and its Applications*, I, John Wiley, 1968.
- Ferguson GA, *Statistical Analysis in Psychology and Education*, 5th ed., McGraw-Hill, 1981.
- Finney DJ, *Statistical Method in Biological Assay*, 3rd ed., Charles Griffin, 1978.
- Fisher RA, *Statistical Methods for Research Workers*, 14th ed., Hafner. 1970.
- Fisher RA and Yates F, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed., Longman, 1974.
- Gabriel KR, 'A procedure for testing the homogeneity of all sets of means in analysis of variance', *Biometrics*, 20 : 459, 1964.
- Garrett HE, *Statistics in Psychology and Education*, Longman, 1966.
- Gibbons JD, *Nonparametric Methods for Quantitative Analysis*, Holt, Rinehart and Winston, 1976.
- Glass GV and Stanley JC, *Statistical Methods in Education and Psychology*, Prentice-Hall, 1970.
- Guenther WC, *The Analysis of Variance*, Prentice-Hall, 1964.
- Guilford JP, *Psychometric Methods*, Tata McGraw-Hill, 1971.
- Guilford JP and Fruchter B, *Fundamental Statistics in Psychology and Education*, 6th ed., McGraw-Hill, 1978.

- Hannan EJ et al (eds.), *Handbook of Statistics*, North-Holland, 1985.
- Harris RJ, *A Primer of Multivariate Statistics*, Academic, 1975.
- Helwig JT, *SAS Introductory Guide*, Statistical Analysis System Institute, 1978.
- Hill AB, *A Short Textbook of Medical Statistics*, 11th ed., Hodder and Stoughton, 1984.
- Hollander M and Wolfe DA, *Nonparametric Statistical Method*, John Wiley, 1973.
- Kendall MG, *Rank Correlation Methods*, 4th ed., Charles Griffin, 1970.
- Keppel G, *Design and Analysis : A Researcher's Handbook*, Prentice-Hall, 1973.
- Kerlinger FM, *Foundations of Behavioral Research*, Holt, Rinehart and Winston, 1986.
- Lindgren BW, *Basic Ideas of Statistics*, Macmillan, 1975.
- Marascuilo L, *Statistical Methods for Behavioral Science Research*, McGraw-Hill, 1971.
- McCormick EJ and Daniel RI, *Industrial Psychology*, 7th ed., Prentice-Hall of India, 1984.
- McGuigan FJ, *Experimental Psychology : Methods of Research*, 5th ed., Prentice-Hall International, 1990.
- Meredith WB, *Basic Mathematical and Statistical Tables for Psychology and Education*, McGraw-Hill, 1967.
- Mosiman JE, *Elements of Probability for the Biological Sciences*, Appleton Century-Crofts, 1968.
- Myers JL, *Fundamentals of Experimental Design*, 3rd ed., Allyn and Bacon, 1979.
- Nie NH et al, *Statistical Package for the Social Sciences*, 2nd ed., McGraw-Hill, 1975.
- Nunnally JC, *Introductory Statistics for Psychology and Education*, McGraw-Hill, 1975.
- Nunnally JC, *Psychometric Theory*, 2nd ed., McGraw-Hill, 1978.
- Scheffe H, *The Analysis of Variance*, John Wiley, 1959.
- Snedecore GW and Cochran WG, *Statistical Methods*, 6th ed., Iowa State University Press, 1967.
- Sokal RR and Rohlf FJ, *Biometry*, Freeman, 1969.
- Sokal RR and Rohlf FJ, *Introduction to Biostatistics*, Freeman, 1973.
- Som RK, *A Manual of Sampling Techniques*, Heinemann, 1973.
- Sprent P, *Applied Nonparametric Statistical Methods*, Chapman and Hall, 1989.
- Sprott DA, *Statistical Inference in Science*, Springer-Verlag, 2000.
- Thorndike RM, *Correlation Procedures For Research*, Gardener, 1978.
- Tufte ER, *The Visual Display of Quantitative Information*, Graphics Press, 1983.
- Weber JC and Lamb DR, *Statistics and Research in Physical Education*, Mosby, 1970.
- Wert JE, *Educational Statistics*, McGraw-Hill.
- Wert JE et al, *Statistical Methods in Education and Psychological Research*, Appleton-Century Crofts, 1954.
- Wilcox RR, *Fundamentals of Modern Statistical Methods*, Springer-Verlag, 2000.
- Wilcoxon F and Wilcox RA, *Some Rapid Approximate Statistical Procedures*, American Cyanamid Co 1964.
- Winer BJ, *Statistical Principles in Experimental Design*, 2nd ed., McGraw-Hill, 1971.

SAMPLE QUESTIONS

CHAPTER 1

1. Complete each of the following statements by choosing and marking with a tick (✓) the correct alternative.

(a) Ratio scale is used to express the values of :

(i) body temperature, (ii) intelligence, (iii) femur length.

(b) Stratified random sampling is used for drawing a sample from a population which is :

(i) small and homogeneous, (ii) large and heterogeneous, (iii) widespread and vast.

(c) With respect to a nominal variable, individuals of a population can be subjected to :

(i) serial gradations into ranks, (ii) assessments of only qualitative differences, (iii) quantitative measurements.

(d) Of the following, the only discrete variable is :

(i) blood volume, (ii) body weight, (iii) respiratory rate.

(e) Interval scale has to be used to express the values of :

(i) heart rate, (ii) body temperature, (iii) cell count.

(f) Multistage sampling is used where the population is :

(i) vast and widespread, (ii) small and homogeneous, (iii) large and heterogeneous.

(g) Of the following, the only continuous variable is :

(i) interorbital width, (ii) sex, (iii) ferocity.

2. Match each item of Column 1 with the correct item in Column 2, mentioning the serial number of the latter in the space after the former.

Column 1

- (a) ordinal variable _____
- (b) absolute measure of dispersion _____
- (c) discrete variable _____
- (d) statistics of location _____
- (e) prediction statistic _____
- (f) nominal variable _____
- (g) sampling statistic _____
- (h) continuous variable _____

Column 2

- (i) median
- (ii) regression coefficient
- (iii) standard error
- (iv) coefficient of variation
- (v) litter size
- (vi) body height
- (vii) variance
- (viii) sex
- (ix) personality

3. Fill up the blanks in the following statements with the correct words chosen from among those given within the parentheses below :

- (a) Coefficient of variation is a _____ measure of dispersion while an independent variable strictly controlled by the investigator is a _____ variable.
- (b) An experiment studies the effects of the _____ variable on a _____ variable.
- (c) Quartile deviation is the _____ measure of dispersion while a _____ variable is an uncontrolled independent variable in an experiment.
- (d) Simple random sampling is applied on a _____ population while quota sampling is done with a _____ population.
- (e) A summary value of the scores of a variable for a population is a _____ while that for a sample is a _____.
- (f) Fixed interval sampling is based on _____ while judgment sampling depends on the _____ by the investigator.
- (stratified, absolute, homogeneous, treatment, relative, absolute, dependent, statistic, parameter, choice, classification, independent, probabilities)
4. Mark the odd item in each of the following series with a tick (✓) mark.
- (a) : (i) blood volume, (ii) blood group, (iii) blood sugar, (iv) blood pressure.
- (b) : (i) multistage sampling, (ii) stratified random sampling ; (iii) fixed interval sampling, (iv) judgement sampling.
- (c) : (i) standard error, (ii) mean, (iii) standard deviation, (iv) correlation coefficient.
- (d) : (i) standard deviation, (ii) mean deviation, (iii) percentile, (iv) range.
5. (a) Explain what you mean by statistic and parameter.
- (b) Classify statistics, defining each class with suitable examples.
- (c) What do you mean by a point estimate and an interval estimate of a parameter ?
6. (a) What is the difference between probability sampling and judgement sampling ?
- (b) Discuss how and when the following methods of sampling are undertaken : simple random sampling, multistage sampling and stratified random sampling.
- (c) What is incidental sampling ?
7. (a) What are the basic characteristics of measurement variables ?
- (b) Classify measurement variables, describing each class briefly with examples.
- (c) Define derived variables with examples.
8. (a) Explain what you mean by dependent, independent and relevant variables with respect to experiments.
- (b) Classify independent variables of experiments, describing each class with suitable examples.
- (c) Describe different classes of relevant variables with examples.
9. (a) Explain the terms population and sample.
- (b) Explain why a sample should be used, instead of the entire population, for an experiment.
- (c) Write briefly about random samplings, respectively with and without replacement.

CHAPTER 2

1. Complete each of the following statements by choosing and marking with a tick (✓) the correct alternative.

(a) Ogives are graphical representations of :

(i) bivariate frequency distributions, (ii) qualitative frequency distributions, (iii) cumulative frequency distributions.

(b) Frequency polygons are area diagrams of :

(i) discrete frequency distributions, (ii) continuous frequency distributions, (iii) qualitative frequency distributions.

(c) The frequency distribution of a discrete variable may be diagrammatically represented by :

(i) histogram, (ii) frequency polygon, (iii) bar diagram.

(d) Frequency distributions of nominal variables can be represented by :

(i) frequency polygon, (ii) pie diagram, (iii) histogram.

(e) True class limits are shown in the frequency distributions of :

(i) continuous variables, (ii) nominal variables, (iii) discontinuous variables.

(f) All individuals in a class interval of a quantitative frequency distribution are assumed to possess the score identical with its :

(i) upper true limit (ii) lower true limit, (iii) midpoint.

(g) the less-than *cf* of an interval of a distribution is the sum of the frequencies from the lowest interval to the

(i) midpoint, (ii) upper true limit, (iii) lower true limit of the relevant interval.

2. Match each item of Column 1 with the correct item in Column 2, putting the serial number of the latter in the space after the former.

Column 1

(a) scattergram _____

(b) *cf* ogive _____

(c) point distribution _____

(d) histogram _____

(e) proportional bar diagram _____

Column 2

(i) cumulative frequency

(ii) discrete variable

(iii) cumulative percentage

(iv) bivariate distribution

(v) continuous variable

(vi) nominal variable

3. Fill up the blanks in the following statements with the correct words chosen from among those given within the parentheses below :

(a) The form of association between two _____ variables is graphically indicated by their _____.

(b) The frequency polygon of a _____ frequency distribution is obtained by plotting the frequency of each class interval against the _____ of that interval.

(c) An ogive may be drawn by plotting the _____ of each interval against its _____.

(d) A pie diagram loses precision if there are too _____ classes in the distribution while smoothening of a frequency polygon decreases its _____.

(jaggedness, scattergram, *cf*, few, midpoint, X_u , measurement, many, continuous)

4. Mark the odd item in each of the following series with a tick (✓) mark.
- (a) : (i) frequency table ; (ii) frequency distribution ; (iii) continuous variable ; (iv) frequency polygon.
- (b) : (i) midpoint ; (ii) continuous distribution ; (iii) true class limits ; (iv) point distribution.
- (c) : (i) cumulative frequencies ; (ii) pie diagram ; (iii) ogive ; (iv) cumulative percentages.
- (d) : (i) simple bar diagram ; (ii) pie diagram ; (iii) histogram ; (iv) proportional bar diagram.
5. (a) Describe the working out of a frequency distribution from the raw scores of a continuous variable.
- (b) Write how a frequency polygon is plotted from a continuous frequency distribution.
- (c) Discuss the advantages and disadvantages of using the frequency polygon.
6. (a) Describe the working out of the cumulative percentage up to the true upper limit of each class interval from the grouped data of a continuous frequency distribution.
- (b) How would you draw an ogive using the cumulative percentages thus computed ?
- (c) Using the following frequency distribution of winglengths (mm) of a sample of cockroaches, work out the cumulative percentages upto the true upper limits of the respective class intervals.
- | Class intervals
(score limits) | 22-25 | 26-29 | 30-33 | 34-37 | 38-41 |
|-----------------------------------|-------|-------|-------|-------|-------|
| Frequencies | 5 | 10 | 20 | 9 | 6 |
7. (a) Describe how you would draw a histogram for the grouped data of a frequency distribution.
- (b) Mention the merits and demerits of a histogram for the graphical representation of frequency distributions.
- (c) Draw a frequency polygon using the data given in Question 6(c). How can you work out a "smoothed" polygon from it ?
8. (a) Describe the simple and multiple bar diagrams as the representations of frequency distributions.
- (b) Draw a multiple bar diagram to represent the following frequency distributions of phenotypes in two *Drosophila* samples from two habitats.
- | Phenotypes | grey-body
red-eye | grey-body
scarlet-eye | black-body
red-eye | black-body
scarlet-eye |
|------------|----------------------|--------------------------|-----------------------|---------------------------|
| Sample 1 : | 90 | 28 | 32 | 10 |
| Sample 2 : | 80 | 35 | 35 | 10 |
- (c) Describe the drawing of a pie diagram and its use.
9. (a) Tabulate the following bodyweight (kg) data of a sample of humans into a frequency distribution, having five suitable class intervals.
- 57, 78, 57, 72, 68, 68, 56, 79, 65, 71, 74, 71, 68, 67, 67, 70, 74, 70, 59, 62, 64, 62, 65, 68, 61, 77, 58, 77, 65, 63, 73, 65, 63, 73, 64, 66, 64, 67, 73, 67.
- (b) Draw a histogram for the representation of this frequency distribution.
- (c) Work out the distribution of cumulative frequencies upto the true upper limits of the respective class intervals of the above-mentioned frequency distribution.

CHAPTER 3

1. Complete every following statement by choosing and marking with a tick (✓) mark the correct alternative given below.
- In a distribution which is perfectly symmetrical bilaterally,
 - median and mode are respectively higher and lower than the mean,
 - mean, median and mode are identical,
 - both mode and median are higher than the mean.
 - For an incomplete frequency distribution with open class interval(s),
 - mean cannot be computed though median and mode can still be computed,
 - mean and median cannot be computed but mode can still be worked out,
 - mean and mode cannot be worked out though median can still be computed.
 - P_{12} is a quantile which belongs to the class of :
 - deciles, (ii) quartiles, (iii) percentiles.
 - The ordinate at the median bisects the area of the frequency distribution in two halves that are always :
 - equal in area and symmetrical, (ii) unequal in area and bilaterally asymmetric, (iii) equal in area.
 - The algebraic sum of deviations of all the scores of a sample from its mean amounts to :
 - a positive integer, (ii) zero, (iii) a negative integer.
2. Fill up the blanks in the following statements choosing the correct words from amongst those within the parentheses below.
- In incomplete frequency distributions with open class intervals, you can compute the _____, but not the _____.
 - An asymmetric distribution causes no deflection of the _____, but the maximum deflection of the _____.
 - Below the median lie _____ of the scores of a sample while _____ of the scores lie below the first decile.
 - The fourth quartile of a frequency distribution is that score below which lie _____ of the scores of a sample while below the fourth percentile lies _____ of all the scores.
(all, half, mean, median, mode, one-tenth, mean, one-fourth, 0.04.)
3. (a) Discuss the properties of the mean.
- (b) Describe how you would compute the mean of the scores grouped into a continuous frequency distribution.
- (c) Compute the mean winglength (mm) of a sample of cockroaches using the frequency distribution given in Question 6(c) of Chapter 2.
4. (a) Describe the properties of median.
- (b) Work out the median and the 3rd quartile of the following bodyheight (cm) distribution of a sample of humans.
- | | | | | | | |
|-------------------|---------|---------|---------|---------|---------|---------|
| Class intervals : | 151-155 | 156-160 | 161-165 | 166-170 | 171-175 | 176-180 |
| Frequencies : | 7 | 16 | 28 | 27 | 14 | 8 |

5. (a) Discuss the properties of mode.
- (b) How can you compute the mode of a distribution with unequal class interval lengths, using the mean and the median ?
- (c) Calculate the mean and the mode of the following distribution of achievement test scores in a sample.
- | | | | | | | |
|-------------------|-------|-------|-------|--------|---------|---------|
| Class intervals : | 67-76 | 77-86 | 87-96 | 97-106 | 107-116 | 117-126 |
| Frequencies : | 8 | 13 | 17 | 20 | 14 | 8 |
6. (a) Describe different types of fractiles.
- (b) How would you work out a percentile of a frequency distribution grouped into class intervals ?
- (c) What are percentile ranks ? How would you determine the percentile rank of a score from the cumulative frequencies of the frequency distribution ?
- (d) Find the median of the following blood sugar values (mg dL^{-1}) of a sample of fasting humans : 65, 68, 70, 73, 73, 75, 79, 79, 79, 82, 83, 84, 84, 86, 87.

CHAPTER 4

1. Complete each of the following statements by choosing and marking by a tick (\checkmark) mark the correct alternative given below.
- (a) Quartile deviation is :
- (i) a quantile, (ii) a relative measure of dispersion, (iii) an absolute measure of dispersion.
- (b) A relative measure of dispersion :
- (i) does not bear the unit of raw scores, (ii) bears the squared unit of raw scores, (iii) bears the same unit as that of raw scores.
- (c) Addition of a constant number to each score of a sample causes the standard deviation :
- (i) to be increased, (ii) to remain unchanged, (iii) to be decreased.
- (d) Variance is :
- (i) a relative measure of dispersion, (ii) in the same unit as the coefficient of variation, (iii) an absolute measure of dispersion.
- (e) Standard deviation is worked out as the :
- (i) square root of the sum of squares, (ii) square root of the mean square, (iii) square root of the sum of absolute deviations of the scores from the mean.
2. Fill up the blanks in the following statements with correct words chosen from those within the parentheses below.
- (a) The standard deviation of a small ungrouped set of scores is worked out using the — in the denominator while that of a large set of scores may be computed using the — as the denominator.
- (b) Coefficient of quartile deviation is a _____ measure of dispersion while range is one of the _____ measures of dispersion.
- (c) Variance is the _____ central moment about the mean while the mean deviation of scores from the mean is the _____ central moment.

- (d) Coefficient of dispersion exceeds 1 in _____ distributions and is a _____ measure of dispersion.
- (e) Variances bear _____ units of raw scores while coefficients of variation bear _____ units of raw scores. (relative, clumped, first, *df*, absolute, third, no, sample-size, relative, squared, second.)
3. Mark the odd member in each of the following series by a tick (✓) mark.
- (a) : (i) sum of squares, (ii) variance, (iii) quartile deviation, (iv) standard deviation.
- (b) : (i) central moment, (ii) range, (iii) standard deviation, (iv) variance.
- (c) : (i) coefficient of variation, (ii) quartile deviation, (iii) root-mean-square, (iv) mean deviation.
- (d) : (i) standard deviation, (ii) absolute measure of dispersion, (iii) range, (iv) coefficient of dispersion.
4. (a) What is standard deviation ? Discuss its properties.
- (b) Explain what you mean by the unbiased *SD* of small samples, mentioning its computational formulae.
- (c) Work out the unbiased *SD* and variance of the following bodyweight (kg) scores of a sample of humans.
57, 78, 57, 63, 73, 65, 70, 74, 70, 67, 58, 77, 65, 67, 73, 67, 72, 68, 56, 63, 73, 64, 66, 64, 62, 65, 68, 61, 58.
- (d) Also compute the coefficients of variation and dispersion for these scores.
5. (a) What is quartile deviation ? Write how you would compute it using the quartiles, mentioning the computation formula.
- (b) Discuss the properties of quartile deviation, mentioning its relations to the skewness and kurtosis of a distribution.
- (c) Work out the quartile deviation and the coefficient of quartile deviation of the bodyheight (cm) distribution presented in Question 4(b) of Chapter 3.
6. (a) Define variance and coefficient of dispersion.
- (b) Work out the variance and *CD* of the winglength distribution of a sample of cockroaches, presented in Question 6(c) of Chapter 2.
- (c) Compute the variance and the coefficient of variation of the interorbital widths (mm) of the following sample of pigeons.
10.4, 13.0, 12.6, 12.5, 10.3, 11.8, 11.6, 12.4, 10.6, 12.9, 10.7, 12.0, 12.5, 11.0, 11.5, 12.2, 11.7, 10.9, 10.6, 11.5, 11.3, 13.0, 11.7, 10.8, 11.1, 12.3.

CHAPTER 5

1. Complete each of the following statements by choosing and marking with a tick (✓) mark the correct alternative given below.
- (a) The standard error of difference between means is :
(i) a relative measure of dispersion, (ii) a sampling statistic, (iii) an absolute measure of dispersion.
- (b) Means of samples drawn by random sampling from the same population differ from each other due to :
(i) mean deviations, (ii) quartile deviations, (iii) sampling errors.

- (c) A sampling distribution of means results from :
 (i) sampling errors, (ii) coefficients of dispersion, (iii) ranges.
- (d) The degrees of freedom of a statistic depend on :
 (i) sample size, (ii) number of precomputed statistics used in its computation, (iii) both.
- (e) Sampling distributions may be worked out :
 (i) both theoretically and experimentally, (ii) experimentally, (iii) theoretically.
- (f) Linear transformations of raw scores change their distribution with respect to :
 (i) kurtosis, (ii) mean and *SD*, (iii) skewness.
2. Indicate with a tick (✓) mark the odd item in each of the following series.
- (a) : (i) standard error, (ii) standard deviation, (iii) sampling error, (iv) sampling distribution.
- (b) : (i) *z* score, (ii) *T* score, (iii) *C* score, (iv) stanine.
- (c) : (i) statistic, (ii) sampling distribution, (iii) standard error, (iv) parameter.
3. Fill up the blanks in the following statements by choosing the correct words from those within the parentheses below.
- (a) Sampling error is the difference between a _____ and the _____.
- (b) Sampling distributions can be worked out _____ using the laws of probability and _____ using the observed scores of randomly drawn samples.
- (c) Standard error of a statistic is a measure of its dispersion around the _____ in the _____ distribution.
- (d) The *z* score is a _____ transformed _____ score.
- (e) The skewness and kurtosis of a distribution may be changed by _____ transformations of raw scores, but not by their _____ transformations.
 (linear, parameter, sampling, theoretically, linearly, statistic, nonlinear, standard, nonlinearly, parameter, experimentally.)
4. (a) Define and explain the degrees of freedom of a statistic with examples.
 (b) What are the sampling errors of a statistic ? Explain with an example.
 (c) Write briefly about the sampling distributions of statistics and of their differences.
 (d) The standard deviation of memory test scores amounted to 2.35 in a group of 30 students, and to 3.15 in another group of 42 students. Work out the *SE* of the difference between the mean memory test scores.
5. (a) What do you mean by the standard error of a statistic ?
 (b) Define the standard error of the mean and the *SE* of difference between means.
 (c) Describe mentioning the computational formulae, how the *SE* of the mean is worked out for samples drawn from different types of populations by different sampling methods.
 (d) Work out the *SE* of the mean using the following interorbital widths (mm) of a sample of pigeons.
 13.3, 12.7, 11.2, 10.8, 12.6, 10.3, 11.8, 11.7, 12.5, 10.4, 11.1, 12.5, 10.2, 11.5, 13.0, 13.2, 10.5,
 10.7, 11.4, 11.7, 10.0, 11.6, 12.7.
6. (a) Explain what is meant by the standard scores, citing examples.

(b) Describe how the difference between two sample means can be transformed into the standard deviate, mentioning the formulae.

(c) Work out the *SE* of the mean of the following bodyweight (kg.) distribution in a sample of humans.

Class intervals :	45-51	52-58	59-65	66-72	73-79	80-86
Frequencies :	5	15	20	14	8	3

7. (a) Write about the *z* score as a linearly transformed standard score.

(b) Discuss the *SE* of difference between means and its computation from the sample variances.

(c) The standard deviations of kneejerk strength scores were found to be 6.60° in 25 athletes and 5.25° in 16 nonathletes. Work out the *SE* of the difference between the means.

CHAPTER 6

1. Complete each of the following statements by choosing and marking with a tick (✓) mark the correct alternative given below.

(a) Normal distribution is a probability distribution based on :

(i) Poisson equation, (ii) Gosset equation, (iii) Gaussian equation.

(b) Student's *t* distribution is a :

(i) theoretical probability distribution, (ii) discrete probability distribution, (iii) experimental probability distribution.

(c) The probability of random occurrence of any one of a number of alternative and mutually exclusive events is given by the :

(i) multiplication theorem, (ii) binomial theorem, (iii) addition theorem.

(d) The probability of the successive occurrence of a given number of independent events at random is given by the :

(i) Gaussian theorem, (ii) multiplication theorem, (iii) binomial theorem.

(e) The sampling distribution of the means of large samples from a non-normally distributed population with a finite variance is given by the :

(i) sampling theory of means, (ii) central limit theorem, (iii) central theorem of probability.

(f) The unit normal curve is :

(i) platykurtic, (ii) leptokurtic, (iii) mesokurtic.

(g) The Bernoulli distribution gives the probabilities of random occurrences of events of a class of a variable having a population distribution obeying the :

(i) binomial distribution, (ii) *t* distribution, (iii) Poisson distribution.

(h) Student's *t* distribution is :

(i) mesokurtic, (ii) leptokurtic, (iii) platykurtic.

2. Indicate with a tick (✓) mark the odd item in each of the following series :

(a) : (i) Student's *t* distribution, (ii) binomial distribution, (iii) Poisson distribution, (iv) skewed distribution.

(b) : (i) continuous distribution, (ii) normal distribution, (iii) *t* distribution, (iv) binomial distribution.

(c) : (i) Poisson distribution, (ii) unit normal curve, (iii) Student's t distribution, (iv) normal distribution.

(d) : (i) binomial distribution, (ii) dichotomized variable distribution, (iii) Poisson distribution, (iv) t distribution.

(e) : (i) leptokurtic distribution, (ii) t distribution, (iii) Poisson distribution, (iv) normal distribution.

3. Match each item of Column 1 with the correct item in Column 2, putting the serial number of the latter in the space after the former.

Column 1

- (a) small sample _____
- (b) level of significance _____
- (c) rare events _____
- (d) Bowley's coefficient _____
- (e) Bernoulli expansion _____
- (f) mesokurtosis _____

Column 2

- (i) skewness
- (ii) unit normal curve
- (iii) critical z
- (iv) binomial distribution
- (v) t distribution
- (vi) asymptotic
- (vii) Poisson distribution

4. Fill up the blanks in the following statements with the correct words from those given within the parentheses below.

- (a) Normal distributions have _____ skewness while Poisson distributions possess _____ skewness.
- (b) The unit normal curve has its highest ordinate at the _____ and its mean amounts to _____.
- (c) The binomial probability distribution of one of the classes of a dichotomous variable is skewed _____ when the events of that class have a _____ proportion in the population than those of the other class.
- (d) The t distribution can be theoretically worked out using the equation of _____ while the normal distribution can be theoretically plotted using the equation of _____.
- (e) The coefficient of dispersion is _____ than 1.00 in binomial distributions and _____ 1.00 in Poisson distributions.
- (f) Probabilities of events of the rare class of a dichotomous variable form a _____ distribution whose mean is identical with the _____.
- (g) The binomial distribution of one of the classes of a dichotomous variable is _____ skewed when the events of that class have a lower proportion in the population than those of the other class, but is _____ skewed when the two classes have an identical proportion.
- (h) The fractional area beyond the two-tail critical z_α in any one of the tails of the unit normal curve amounts to _____ the corresponding two-tail α while that beyond the one-tail critical z_α in any tail _____ the corresponding one-tail α .

(negatively, variance, equals, no, half, Gauss, mean, positive, not, less, Gossett, positively, higher, Poisson, zero, Bernoulli, equals.)

5. (a) What is the Poisson distribution ? Describe its properties.
- (b) Discuss the assumptions for applying the Poisson distribution to the data.
- (c) How do you use the Poisson distribution to find the probability of random occurrence of X number of cases of the rare class of a dichotomous variable in a sample of size n , drawn from a population known to have the proportion p of such rare cases ?

- (d) Work out and interpret the probability of random occurrence of 4 Down syndrome cases in a sample of 200 humans from a population with 12 such cases on average per 1000 individuals.
6. (a) Compare the binomial and Poisson distributions.
 (b) Discuss the assumptions underlying the use of the binomial distribution.
 (c) Work out and interpret the binomial probability of random occurrence of 6 fluorosis cases in a sample of 20 individuals drawn from a population known to have 30% incidence of fluorosis.
7. (a) Compare the properties of normal and Student's t distributions.
 (b) Write the mathematical equation used in the theoretical computation of normal probability distributions, and its modified form for plotting the unit normal curve.
 (c) Write briefly about the two-tail and one-tail critical z scores.
8. (a) What is the unit normal curve? Describe its principal properties.
 (b) Explain what you mean by the best-fitting normal distribution.
 (c) Work out the best-fitting normal distribution for the following observed frequency distribution of bodyweight scores (kg) of a sample of 120 humans.
- | | | | | | | | |
|-------------------|-------|-------|-------|-------|-------|-------|-------|
| Class intervals : | 41-47 | 48-54 | 55-61 | 62-68 | 69-75 | 76-82 | 83-89 |
| Frequencies : | 5 | 15 | 25 | 43 | 21 | 10 | 1 |
9. (a) Describe the principal properties of Student's t distributions.
 (b) Explain why t distributions are applicable to both small and large samples.
 (c) What are critical t scores?
10. (a) Describe the properties of skewed distributions.
 (b) Mention the formulae for working out different measures of skewness.
 (c) Work out Pearson's second coefficient of skewness of an observed distribution of interorbital widths of a sample of pigeons, having the following statistics.
- Mean : 12.0 mm ; Median : 13.6 mm ; SD : 1.70 mm.
11. (a) Describe different types of kurtosis of distributions with examples.
 (b) Write briefly about the measures of kurtosis with their computational formulae.
 (c) Work out and interpret the percentile coefficient of kurtosis and Bowley's quartile coefficient of skewness of a frequency distribution of achievement test scores, having the following values of its percentiles.
- $P_{10} = 98.8$; $P_{25} = 101.0$; $P_{50} = 118.8$; $P_{75} = 119.7$; $P_{90} = 127.6$.
12. (a) Explain what are confidence intervals and fiducial probabilities.
 (b) Describe with computational formulae the working out of the confidence intervals for means, using z and t scores respectively.
 (c) The mean and SD of winglength scores of 12 houseflies were found to be 4.8 mm and 0.74 mm respectively. Find the confidence limits of the population mean at 0.95 and 0.99 fiducial probabilities, using t scores.
 (d) The mean and SD of numerical operation test scores of 40 girls were found to be 37.5 and 6.35 respectively. Find the 95% confidence limits of the population mean, using z scores.

CHAPTER 7

1. Identify and mark with a tick (✓) mark the correct alternative, given below, for completing each of the following statements.

- (a) Probability of type I error of inference equals that fractional area of the H_0 distributions which equals :
 (i) the rejection region of the H_0 distribution, (ii) the acceptance region of the H_0 distribution, (iii) the area of overlap of the H_a and the H_0 distributions.
- (b) The null hypothesis proposes in any experiment that the observed results have occurred merely due to :
 (i) an association between the dependent and independent variables, (ii) chances of random sampling depending on the laws of probability, (iii) reasons other than these.
- (c) The t tests can be applied if the dependent variable in the experiment happens to be :
 (i) a discrete variable, (ii) a nominal variable, (iii) a continuous variable.
- (d) Probability of type II error of inference equals :
 (i) the rejection region of the H_0 distribution, (ii) the area of overlap of the H_a distribution with the acceptance region of the H_0 distribution, (iii) the acceptance region of the H_0 distribution.
- (e) t tests can be used for the samples drawn from a population where the dependent variable is normally distributed, if the samples are :
 (i) small, (ii) large, (iii) either small or large.
- (f) The unit normal curve and z scores can be used for finding the significance of difference between two group means, provided :
 (i) the groups are small and the variable is normally distributed, (ii) the groups are large and the variable is normally distributed, (iii) the groups are large and the dependent variable is a discrete variable.

2. Match each item of Column 1 with the correct item in Column 2, mentioning the serial number of the latter in the space after the former.

Column 1

Column 2

- | | |
|-----------------------------|------------------------------------|
| (a) rejection region _____ | (i) significant difference |
| (b) one-tail test _____ | (ii) wrongful rejection of H_0 |
| (c) type II error _____ | (iii) random sampling |
| (d) null hypothesis _____ | (iv) critical z score |
| (e) pooled SD _____ | (v) small single-group |
| (f) type I error _____ | (vi) large single-group |
| (g) difference method _____ | (vii) wrongful acceptance of H_0 |
| (h) two-tail test _____ | (viii) significantly higher/lower |
| | (ix) small independent groups |

3. Fill up the blanks in the following statements with the correct words chosen from those given within the parentheses below.

- (a) Probability of type I error of inference would _____ with the _____ in the level of significance.
- (b) The null hypothesis is _____ if the computed z score falls within the _____ region of the H_0 distribution.

- (c) Whether or not one group mean is significantly lower than the other is explored by a _____ test while a _____ test is used to find if one group mean is significantly different from the other.
- (d) Type II error of inference consists of a wrongful _____ of the H_0 while type I error consists of a wrongful _____ of the H_0 .
- (e) The alternative hypothesis is _____ if the computed z score falls within the _____ region of the H_0 distribution.
- (f) Probability of type II error of inference would _____ with the _____ in the level of significance.
- (g) For a test for a significant difference between two means, _____ of the critical region lies in one of the tails of the H_0 distribution while in a test for one mean being significantly higher than the other, one of the tails carries _____ of the critical region.
- (fall, rejection, whole, retained, rise, two-tail, half, accepted, rise, acceptance, rise, rejection, one-tail, rejected, acceptance)
4. (a) Discuss the assumptions underlying the Student's t test.
- (b) Explain what is meant by matched-pair groups.
- (c) Describe mentioning computational formulae how you would work out Student's t score from the dependent variable scores of a large single-group experiment.
- (d) The mean and the unbiased SD of steadiness test scores of a group of 41 mentally retarded children were found to be 70.2 and 9.40, respectively, before any treatment; but the same group had the mean and the SD of 79.6 and 11.2 respectively, after a course of treatment. The product-moment r between their test scores, respectively before and after the treatment, was worked out to be +0.55. Find whether or not the mean steadiness score was significantly higher after the treatment than before it.
5. (a) Describe one-tail and two-tail tests with examples.
- (b) Compare the acceptance and rejection regions of the H_0 distribution for a two-tail test with respectively those for a one-tail test.
- (c) Work out an appropriate t test to find whether or not the mean score is significantly higher in the second trial than in the first trial in a sensory-motor test administered to the following group of 10 subjects.
- | Individual : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------------|----|----|----|----|----|----|----|----|----|----|
| Test scores : | | | | | | | | | | |
| first trial : | 50 | 62 | 40 | 55 | 62 | 41 | 28 | 37 | 35 | 46 |
| second trial : | 62 | 65 | 54 | 53 | 74 | 55 | 38 | 48 | 44 | 63 |
6. (a) What is the level of significance? Discuss its relation with the errors of inference.
- (b) What are the assumptions for using the z score in significance tests?
- (c) The mean and SD of birthweights were respectively 3.2 kg and 0.65 kg for 115 first-born infants, but amounted to respectively 2.4 kg and 0.52 kg for 105 third-born infants. (i) Find if the mean birthweights differ significantly in the two groups, using the unit normal curve areas. (ii) Also find whether or not the mean birthweight is significantly higher in first-born infants than in third-born infants.
7. (a) Explain what is meant by the null hypothesis and the alternative hypothesis, and how they are used in finding the significance of difference between means.
- (b) Describe how Student's t score is worked out and tested for the significance of difference between means of two sets of scores in a small single-group experiment.

- (c) Apply t test to find whether or not there is a significant difference between the mean grip strengths (kg) of the following group of 10 persons in resting and fatigued conditions.

Individuals :	1	2	3	4	5	6	7	8	9	10
Grip strengths :										
resting :	10	12	14	9	11	13	10	8	12	15
fatigued :	4	6	7	5	8	11	5	4	9	8

8. (a) Give an account of the assumptions for t test.
 (b) Describe mentioning the computational formulae how you would work out the t test for significance of difference between the means of (i) large unequal-size independent groups and (ii) small unequal-size independent groups.
 (c) Use t test to find whether or not the mean grip strength of the following group of athletes is significantly higher than that of the nonathletes.

Athletes :	10,	12,	9,	16,	11,	14,	15,	13,	11,	12,	14.
Nonathletes :	6,	4,	3,	7,	6,	5,	8,	4,	9.		

9. (a) Describe the importance of the df of t in finding its significance, and how the df may be worked out for unequal-size and equal-size independent groups and for single group experiments.
 (b) Compare mentioning the formulae the working out of the SE of difference between means from the dependent variable scores of small unequal-size independent groups and of large single-group experiments.
 (c) Apply an appropriate t test to find whether or not there is a significant difference between the mean interorbital widths (mm) of the following groups of male and female pigeons.

Males :	11.3,	12.2,	12.8,	11.9,	13.0,	13.4,	11.8,	12.7,	12.9,	13.3,	12.4.
Females :	10.5,	10.0,	10.4,	11.0,	10.9,	10.7,	11.3,	10.8,	10.2,	11.5.	

CHAPTER 8

1. Match each item of Column 1 with the correct item of Column 2 and put the serial number of the latter in the space against the former.

Column 1

- (a) multiple regression _____
 (b) phi coefficient _____
 (c) Spearman's rho _____
 (d) multiple correlation _____
 (e) slope of regression line _____
 (f) model I regression _____
 (g) biserial r _____
 (h) model II regression _____
 (i) product-moment r _____

Column 2

- (i) average ranks
 (ii) Fisher's z transformation
 (iii) regression coefficient
 (iv) sum of quotients
 (v) genuinely dichotomous variables
 (vi) classification variable
 (vii) artificially dichotomized variable
 (viii) treatment variable
 (ix) partial regression coefficients.
 (x) variance ratio.

2. Complete each of the following statements by choosing and marking by a tick (✓) mark the correct alternative out of those given below.
- Variables with more than two classes are correlated by :
(i) tetrachoric r , (ii) biserial r , (iii) phi coefficient, (iv) contingency coefficient.
 - If the parametric correlation coefficient amounts to + 0.37, the sampling distribution of product-moment r values has :
(i) no skewness, (ii) negative skewness, (iii) positive skewness, (iv) either (ii) or (iii).
 - Two ordinal variables are correlated by :
(i) tetrachoric r , (ii) Kendall's tau, (iii) phi coefficient, (iv) contingency coefficient.
 - Linear regression can be worked out only if both dependent and independent variables have :
(i) discrete distributions, (ii) significant linear correlation, (iii) continuous distributions, (iv) both (ii) and (iii).
 - A normally distributed continuous variable is artificially dichotomized if in the sample its scores are :
(i) too many, (ii) non-normally distributed, (iii) distributed in a truncated form, (iv) either (ii) or (iii).
 - A genuinely dichotomous variable is correlated with a continuous measurement variable by :
(i) point-biserial r , (ii) biserial r , (iii) phi coefficient, (iv) Spearman's rho.
3. Fill up the blanks in each of the following statements by correct words chosen from those given within the parentheses below.
- The significance of the computed multiple correlation coefficient may be tested by converting it either to _____ score or to _____ ratio.
 - The slope of the regression line of a criterion on one of the predictors is given by a _____ coefficient when the other predictors are partialled out.
 - Deviations of actual criterion scores from the predicted criterion score in a regression are estimated by the _____ of _____.
 - Phi coefficient correlates two _____ variables while Kendall's tau correlates two _____ variables.
 - Product-moment r has _____ skewed sampling distributions if the parametric ρ has a positive value, while its sampling distributions are _____ skewed if ρ has a zero value.
 - Giving average ranks to tied scores is a source of error for two correlation statistics, _____ and _____ (rho, regression, negatively, SE, variance, not, partial, ordinal, t , tau, dichotomous, continuous, estimate)
4. Mark the odd item in each of the following series with a tick (✓) mark.
- (i) biserial r , (ii) Spearman's rho, (iii) phi coefficient, (iv) point biserial r .
 - (i) multiple correlation, (ii) product-moment correlation, (iii) Kendall's tau, (iv) Spearman's rho.
 - (i) partial correlation, (ii) multiple correlation, (iii) multiple regression, (iv) simple regression.
 - (i) b_{YX} , (ii) $b_{12.3}$, (iii) β_3 , (iv) $b_{13.2}$.
5. (a) Discuss the assumptions underlying the product-moment r .
- (b) Give different formulae for working out product-moment r respectively from raw scores, sum of products, covariance and sums of squares of ungrouped data. How do you test its significance ?

- (c) Work out the product-moment r between the tracheal ventilation scores (ml/min) and the oxygen consumption scores (ml/min) of the following sample of locusts, and test its significance.

Individual	:	1	2	3	4	5	6	7	8	9	10
Ventilation	:	85	80	70	68	68	75	70	60	71	73
O ₂ consumption	:	4.0	3.4	2.5	2.7	2.5	3.2	3.0	2.5	3.0	3.2

6. (a) What are the assumptions for Spearman's rank-difference correlation coefficient ?
 (b) Describe, mentioning the computational formulae, the working out of Spearman's rho and the testing of its significance.
 (c) Work out Spearman's rho with the data of the preceding Q.5(c) and test its significance.
7. (a) What is partial correlation ? Describe its assumptions.
 (b) Describe the computation of the first-order partial r between variables X_1 and X_2 , partialling out variable X_3 , and the testing of significance of the computed partial r .
 (c) Use the following data from a sample of 53 humans to find whether or not there is a significant partial linear correlation between glomerular filtration rate (X_1 ml/min) and glomerular blood pressure (X_2 mm Hg) when the effect of plasma protein osmotic pressure (X_3 mm Hg) is partialled out.
- $$r_{12} = +0.68 ; \quad r_{13} = -0.32 ; \quad r_{23} = +0.18.$$
- Also find if there is a significant multiple linear correlation between X_1 and the combination of X_2 and X_3 in this case.
8. (a) Describe the assumptions for multiple linear correlation.
 (b) Describe, mentioning computational formulae, how you work out the multiple linear correlation between a criterion (X_1) and the weighted sum of two predictors (X_2 and X_3), and test its significance using critical t and critical F values.
 (c) Using the following data from a sample of 33 students, find whether or not there is a significant first-order partial correlation between their mathematical aptitude test scores (X_1) and abstract reasoning test scores (X_2), partialling out their numerical test scores (X_3).
- $$r_{12} = +0.53 ; \quad r_{13} = +0.22 ; \quad r_{23} = +0.13.$$
- Also find if there is a significant multiple linear correlation between X_1 and the combination of X_2 and X_3 , using critical F values.
9. (a) Describe the assumptions for Kendall's tau.
 (b) How do you work out Kendall's tau for correlating the scores of two continuous variables, and find its significance ?
 (c) Work out Kendall's tau between the gill weights (mg) and body weights (g) of a sample of 10 crabs, given in Example 8.2.4. (page 146) to find whether or not there is a significant correlation between them.
10. (a) Describe the properties of product-moment r .
 (b) How do you test the significance of the computed r , by transforming it respectively to Student's t and Fisher's z ?
 (c) Work out product-moment r between the vocabulary test scores of the following students and the marks obtained by them in English in a school examination, and test its significance.

Student	:	1	2	3	4	5	6	7	8	9	10
Vocabulary score	:	12	36	10	25	34	23	9	30	11	30
English marks	:	37	59	31	50	60	49	28	54	45	58

11. (a) Why would you work out point biserial r ? What are the assumptions for point biserial r ?
 (b) Describe mentioning formulae how to work out r_{pbi} and test its significance.
 (c) Work out a suitable correlation coefficient with the following pulmonary tidal volume scores (ml) of 10 males and 10 females to find whether or not there is a significant correlation between sex and tidal volume. Justify your choice of the correlation coefficient.
- Males : 520, 492, 456, 525, 515, 550, 490, 520, 545, 500.
 Females : 460, 476, 440, 410, 404, 385, 370, 402, 375, 355.
12. (a) In what cases is a continuous measurement variable dichotomized for linear correlation?
 (b) Describe with computational formulae the computation of biserial r and the test of its significance. How are r_b and r_{pbi} interconverted to one another?
 (c) Find whether or not there is a significant linear correlation between systolic blood pressure (mm Hg) and diabetes, using the following systolic BP data of 10 diabetics and 10 nondiabetics and justifying your choice of the correlation coefficient.
- Diabetics : 244, 158, 140, 208, 180, 216, 162, 204, 232, 174.
 Nondiabetics : 120, 142, 132, 138, 160, 118, 140, 150, 122, 130.
13. (a) Write about different models of regression with examples.
 (b) Discuss the assumptions for simple linear regression.
 (c) Use the data of the preceding Q.5(c) to work out the linear regression equation of O_2 consumption on tracheal ventilation in locusts.
14. (a) Describe the properties of simple linear regression.
 (b) Write mentioning computational formulae how you would work out the linear regression of variable Y on variable X , using respectively the raw scores, the sum of products, the covariance, and the product-moment r .
 (c) Use the data of the preceding Q.10(c) to compute the linear regression of the examination marks in English on the vocabulary scores of students.
15. (a) What is multiple linear regression? Discuss its assumptions.
 (b) Describe mentioning formulae how you compute the multiple linear regression of a criterion X_1 on the given values of predictors X_2 and X_3 , and the SE of estimate of the predicted values of the criterion.
 (c) Use the data of the preceding Q.7(c) to work out the linear regression of glomerular filtration rate on glomerular blood pressure and plasma protein osmotic pressure. Also compute the coefficient of multiple determination and the SE of estimate.
16. (a) What is the purpose of using the contingency coefficient? Comment on the assumptions needed for its application.
 (b) Describe the computation of contingency coefficient and the test for its significance.
 (c) Work out the contingency coefficient to find if there is a significant correlation between the socioeconomic

classes of students and the undergraduate courses of their enrolment, using the following distribution of students from different socioeconomic classes enrolled in different courses.

Courses	Socioeconomic classes				Total
	1	2	3	4	
Humanities	10	72	95	15	192
Science	25	80	125	95	325
Commerce	22	40	86	55	203
Total	57	192	306	165	720

17. (a) Explain where you would use biserial r for correlation, with examples.
 (b) Discuss the assumptions underlying the use of biserial r .
 (c) Describe the computation and the test for significance of biserial r .
 (d) Find whether or not there is a significant correlation between two given IQ groups of students and their memory test scores, using the following data and justifying your choice of the correlation coefficient.

Memory test score ranges :	40-44	45-49	50-54	55-59	60-64	65-69	70-74
No. of students :							
with IQ < 120 :	15	22	28	14	11	8	2
with IQ \geq 120 :	2	4	8	25	32	17	12

18. (a) Compare the applications and assumptions of point biserial r with those of biserial r .
 (b) Discuss when you would dichotomize the scores of a continuous variable to form an artificially dichotomous variable.
 (c) How can you convert r_b into r_{pbi} and vice versa ?
 (d) Find whether or not there is a significant correlation between sex and total memory test scores, using the following data and justifying your choice of the correlation coefficient.

Memory test score ranges :	41-45	46-50	51-55	56-60	61-65	66-70	71-75
No. of females :	3	3	9	22	35	20	8
No. of males :	2	4	12	25	30	16	11

CHAPTER 9

1. Mark the odd item in each of the following series with a tick (\checkmark) mark.
 (a) : (i) Yates' correction ; (ii) chi square ; (iii) rank sum ; (iv) G test.
 (b) : (i) goodness of fit ; (ii) contingency table ; (iii) association ; (iv) independence.
 (c) : (i) analysis of frequencies ; (ii) goodness of fit ; (iii) independence ; (iv) signed rank.
 (d) : (i) composite rank test ; (ii) chi square test ; (iii) Mann-Whitney test ; (iv) median test.
2. Match each item of Column 1 with the correct item in Column 2 and put the serial number of the latter in the space after the former.

Column 1

- (a) signed ranks _____
- (b) Yates' correction _____
- (c) composite ranks _____
- (d) contingency table _____
- (e) statistic G _____

Column 2

- (i) association
- (ii) t test
- (iii) chi square
- (iv) paired observations
- (v) independent groups
- (vi) log-likelihood ratio statistic

3. Fill up the blanks in each of the following statements with the correct words chosen from those within the parentheses below.
- (a) Whether or not an observed distribution conforms significantly to a proposed distribution may be tested by _____ test or _____ test.
 - (b) The Mann-Whitney test is a nonparametric alternative to the t test for _____ groups of _____ sizes.
 - (c) The significance of difference between more than two group means can be tested nonparametrically by the _____ test, based on the H_0 proposing a common _____ of all the groups.
 - (d) The chi square test of _____ requires the framing of a _____ table showing the association between variables.
 - (e) Wilcoxon's _____ rank test and _____ rank test are the nonparametric alternatives to t tests for respectively independent and single-group experiments.
 - (f) Each negative $(f_o - f_e)$ is _____ by 0.5 and each positive $(f_o - f_e)$ is _____ by 0.5 for Yates' correction in chi square tests.
 - (g) The G test for goodness of fit needs no Yates' correction so long as the sample size _____ 200 and there are _____ than two classes in the observed distribution.
(exceeds, contingency, independent, more, decreased, equal, G , median, signed, increased, chi-square, independence, unequal, composite, median)
4. Complete each of the following statements by choosing and marking with a tick (\checkmark) mark the correct alternative given below.
- (a) Whether or not an observed distribution conforms to a normal distribution can be tested by :
(i) chi square test for goodness of fit, (ii) Chi square test of independence, (iii) G test for goodness of fit, (iv) either (i) or (iii).
 - (b) In two-tail composite rank tests for small unequal-size sample, the sample means do not differ significantly if the smaller sum of ranks :
(i) falls between the critical upper and lower T values, (ii) exceeds the critical upper T value, (iii) is lower than the critical lower T value, (iv) equals the critical lower T .
 - (c) In a two-tail rank sum test for large equal-size groups, the group means differ significantly if the smaller sum of ranks :
(i) exceeds the critical T_α computed from \bar{T} , (ii) does not exceed the critical T_α computed from \bar{T} , (iii) is lower than the critical lower T value of table, (iv) exceeds the critical upper T value of table.
 - (d) For samples not below 8 in size, the Mann-Whitney U is computed from :
(i) the bigger sum of ranks, (ii) the smaller sum of ranks, (iii) any of the two rank sums, (iv) both the rank sums.

- (e) A fourfold contingency table is used for computing chi square if :
- (i) the variable in a test for goodness of fit is divided into four classes, (ii) the variables in a test of independence are each divided into two classes, (iii) the variable in a test for goodness of fit is divided into two classes, (iv) the variables in a test of independence are each divided into four classes.

5. (a) Describe the working out of a nonparametric statistic proposed by Wilcoxon for the significance of difference between the unpaired observations of two equal-size independent groups. What are the sources of its inaccuracies ?

- (b) Elaborate on how you would test the significance of such computed statistic in case of (i) small unequal-size groups and (ii) large equal-size groups.

- (c) Apply an appropriate nonparametric test to find whether or not the mean winglengths (mm) of the following two groups of cockroaches differ significantly. Justify your choice of the test.

Group I : 28, 34, 37, 35, 42, 29, 38, 46, 40, 39.

Group II: 24, 20, 23, 22, 28, 20, 21, 23, 21, 28.

6. (a) Describe the computation and interpretation of a powerful nonparametric statistic preferable for testing the significance of difference between unpaired observations of two unequal-size independent groups, mentioning the differences in the method according to the group sizes.

- (b) Work out and interpret an appropriate nonparametric statistic for finding any significant difference between the mean grip strengths (kg) of the following two independent groups of athletes, justifying your choice of the statistic.

Group I : 10, 12, 14, 9, 11, 13, 10, 8, 12, 15, 14.

Group II : 14, 6, 7, 5, 8, 11, 5, 4, 9, 8.

- (c) Use the same data to work out and interpret an appropriate Wilcoxon test.

7. (a) Name and describe a nonparametric test of Wilcoxon for the significance of difference between paired observations of single-group experiments, mentioning the difference in the method according to the sample size.

- (b) Work out the afore-said statistic to find whether or not the mean memory test scores, before and after practice, differ significantly in the following group of subjects.

Subject :	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Before :	21	18	17	19	35	32	13	12	27	30	20	23	20	8	31	16
After :	26	39	35	38	28	24	30	28	19	37	25	32	20	13	31	33
Subject :	17	18	19	20	21	22	23	24	25	26	27	28				
Before :	11	35	19	28	23	18	30	29	23	29	32	30				
After :	32	24	37	38	27	22	32	16	39	35	39	32				

- (c) Apply Wilcoxon's composite rank test to the following winglengths (mm) of two groups of houseflies to find if there is a significant difference between the group means.

Group I : 4.0 4.8, 3.9, 5.3, 5.4, 5.5 4.7, 5.4.

Group II : 3.2, 3.6, 3.9, 3.3, 3.9, 4.0, 3.5.

8. (a) Describe where and how you would work out the median test and interpret the result.

- (b) When would you use the fourfold contingency table for this test.

- (c) Apply the median test to the following tracheal ventilation scores (ml/min) of three samples of beetles, exposed to three different levels of a pesticide, to find if there are significant differences between the scores of different groups.

Group I	:	80.6, 85.0, 82.7, 78.7, 84.6, 85.0, 87.0, 86.4, 80.7.
Group II	:	71.7, 77.2, 70.3, 69.2, 67.3, 70.1, 67.8, 73.7.
Group III	:	58.5, 65.0, 59.1, 56.7, 57.6, 59.0, 61.2, 59.6, 62.7.

9. (a) What are nonparametric statistics ? Compare them with parametric statistics, with examples.
 (b) Why is the nonparametric chi square test considered as an analysis of frequencies ? Mention its basic formula and properties.
 (c) Describe the probability distributions of chi squares. How do you determine the *df* of chi square for the tests of goodness of fit with normal, binomial and Mendelian phenotype distributions, respectively ?
 (d) Use the chi square test to find whether or not there is a significant goodness of fit between the following phenotype distribution of *Drosophila* and the Mendelian 9:3:3:1 distribution.

Phenotypes :	Grey-bodied red-eyed (AB)	Grey-bodied scarlet-eyed (Ab)	Black-bodied red-eyed (aB)	Black-bodied scarlet-eyed (ab)
No. of flies :	105	32	37	18

10. (a) On sampling 80 litters of an animal species having a population male : female ratio of 0.52 : 0.48, the litter-wise frequency distribution of females was found to be as follows. Use the chi square test to find whether or not the observed litterwise distribution of females has a significant goodness of fit with the binomial distribution expected from the given population sex-ratio.

No. of females per litter:	0	1	2	3	4	5
No. of litters	: 2	15	26	22	12	3

- (b) What is Yates' correction ? Describe when and how you would apply it in a chi square test for goodness of fit.
 (c) Apply chi square test to find whether there is a significant difference between the following observed frequency distribution (f_o) of body heights (cm) of 90 humans and the best-fitting normal distribution (f_e).

Classes :	146-150	151-155	156-160	161-165	166-170	171-175	176-180.
f_o :	4	12	12	24	20	15	3
f_e :	4.3	9.9	17.1	21.4	19.0	12.3	6.0

11. (a) Explain what is meant by the chi square test of independence.
 (b) Describe, mentioning formulae, how you would frame a contingency table and use it for the chi square test of independence.
 (c) Describe the use of a fourfold contingency table for computing chi square in an alternative way, without working out the expected frequencies of cells.
 (d) Out of 60 subjects passing in a psychological test, respectively 30, 17 and 13 individuals were rated subsequently as above-average, average and below-average in laboratory performance. But out of 30 subjects failing in the psychological test, 4, 11 and 15 individuals got respectively the above-average, average and below-average ratings in laboratory performance. Use chi square test to find whether or not laboratory performance is independent of the psychological test results.

12. (a) Describe the working out of the G test for goodness of fit, mentioning the computational formula and the determination of the df according to the nature of the proposed distribution.
- (b) When and how you would apply Yates' correction in the G test ?
- (c) Use the G test to find whether or not there is a significant goodness of fit between the following phenotype distribution of *Drosophila* and the Mendelian 9:3:3:1 distribution.

Phenotypes :	Grey-red (AB)	Grey-scarlet (Ab)	Black-red (aB)	Black-scarlet (ab)
No. of flies :	118	38	36	16

13. (a) What is a G test of independence ? Describe how it is worked out mentioning the computational formula and the df .
- (b) When and how you would use Yates' correction in such a test ?
- (c) Solve the psychological problem presented in the preceding Q.11(d), using the G test instead of the chi square test.
14. (a) Out of 80 diabetics, 32 were found suffering from hypercholesterolemia while the rest had normal serum cholesterol. Out of 70 nondiabetics, only 14 were hypercholesterolemic. Use chi square test of independence to find if there is any significant association between diabetes and hypercholesterolemia.
- (b) Out of 45 hypertensive humans, 28 were found to be hyper-reactors to cold, showing a rise of more than 20 mm Hg in their diastolic pressure on exposure to cold ; but the remaining 17 were normoreactors with less than 20 mm Hg rise in pressure on cold exposure. Out of 35 non-hypertensive humans, only 7 were hyper-reactors to cold while others were normoreactors. Apply G test of independence to find whether or not there is a significant association between hypertension and hyper-reaction to cold.

CHAPTER 10

1. Fill up the blanks in the following statements with the correct words chosen from those given within the parentheses below.
- (a) Validity of a test is affected mostly by _____ errors of measurement while reliability is affected mostly by _____ errors of measurement.
- (b) Reliability coefficient is that proportion of the _____ variance of test scores which is its _____ variance.
- (c) Reliability coefficient is a measure of _____ of a test while the index of reliability is a measure of _____ correlation.
- (d) The consistency of results of a test on repeated application on the same group is given by its _____ while the capacity of a test to measure the specific variable in exclusion of others is given by its _____.
- (e) Reliability depends on the proportion of _____ variance in a test while validity depends basically on the _____ variance.
- (f) _____ variables are not deliberately used in an experiment but still affect the dependent variable while _____ variables are deliberately used for studying their effects in the experiment.
- (g) Coefficient of _____ is the reliability coefficient estimated by the _____ method of estimating reliability.
- (h) The capacity of test scores to denote the true scores is given by the _____ validity coefficient which is the square-root of the _____ coefficient of that test.
- (reliability, relevant, true, validity, equivalent, test-whole, independent, random, common-factor, reliability, systematic, self-correlation, alternate-forms, total, intrinsic, true)

2. Match each item of Column 1 with the correct item of Column 2 and put the serial number of the latter in the space after the former.

Column 1

- (a) coefficient of stability _____
 (b) difficulty value _____
 (c) relevant variable _____
 (d) coefficient of internal consistency _____
 (e) expectancy table _____
 (f) coefficient of equivalence _____
 (g) treatment variable _____

Column 2

- (i) parallel-forms
 (ii) extraneous variable
 (iii) independent variable
 (iv) test-retest
 (v) construct validity
 (vi) item analysis
 (vii) rational equivalence
 (viii) scattergram

3. Mark the odd item in each of the following series with a tick (✓) mark.

- (a) : (i) index of discrimination, (ii) coefficient of internal consistency, (iii) index of reliability, (iv) coefficient of equivalence.
 (b) : (i) organismic variables, (ii) treatment variables, (iii) situational relevant variables, (iv) stimulus variables.
 (c) : (i) K-R 20, (ii) Spearman-Brown formula, (iii) σ value of difficulty, (iv) Kuder-Richardson formula 21.
 (d) : (i) predictive validity, (ii) content validity, (iii) concurrent validity, (iv) criterion-related validity.
 (e) : (i) coefficient of internal consistency, (ii) coefficient of stability, (iii) coefficient of test-retest reliability, (iv) coefficient of stability and equivalence.

4. Fill up the blanks in each of the following statements by correct words chosen from those within the parentheses below.

- (a) Reliability of a test is affected mostly by _____ errors of measurement while validity is affected mostly by _____ errors of measurement.
 (b) The coefficient of internal consistency of a test measures both the _____ of its items and _____ of its contents.
 (c) The increase in the number of alternative answers to test items _____ the guess factor while the increase in the number of test items _____ the reliability coefficient.
 (d) Both _____ variables and _____ variables include organismic variables.
 (e) The consistency of results of a test on its repeated application on a sample is given by the _____ of that test while the capacity of a test to measure only a specific variable is given by its _____.
 (f) The _____ validity assesses a test as a measure of the present status of the tested individuals while the _____ validity assesses a test as a measure of their future status with respect to a specific variable.
 (g) Reliability of a test depends on _____ difficulty levels and _____ internal consistency of the test items.
 (h) Predictive validity of a test is dependent upon _____ difficulty levels and _____ correlations of its items. (homogeneity, low, systematic, decreases, identical, random, subject-relevant, concurrent, independent, dependent, equivalence, validity, increases, high, reliability, differing, predictive)

5. Compare each following pair of items.
 - (a) Organismic and stimulus variables.
 - (b) Relevant variables and intervening variables.
 - (c) Power and speed tests.
 - (d) Ratio IQ and deviation IQ.
 - (e) C scale and stanine.
 - (f) Random errors of measurement and systematic errors.
 - (g) T scale and C scale.
 - (h) Correlation table and expectancy table.
 - (i) Linear transformations and nonlinear transformations.
6. (a) Explain the difficulty value and the discriminatory value of test items.
 - (b) Discuss the properties of difficulty value.
 - (c) Describe the methods of estimation of discriminatory value.
7. (a) Describe different classes of independent variables of psychological experiments.
 - (b) Explain what are the relevant variables in psychological experiments and describe their classes with examples.
 - (c) What are intervening variables ?
8. (a) What is meant by the reliability of a test ?
 - (b) Discuss the test-retest method and the split-half method of estimating the reliability of a test.
 - (c) Use the Spearman-Brown formula to work out how long a test, consisting of 22 items and having a reliability coefficient of 0.32, be made to raise its reliability coefficient by 0.18 ?
9. (a) Describe the coefficients for expressing the equivalent-forms reliability.
 - (b) Discuss the alternate-forms method and the rational equivalence method for estimating test reliability.
 - (c) Use the Kuder-Richardson formula 21 to work out the coefficient of internal consistency of a test of 40 items where the mean and the SD of the total scores amount to 28 and 4.5 respectively.
10. (a) Discuss the split-half method and the rational equivalence method for estimating test reliability.
 - (b) Mention the Spearman-Brown formula and the Kuder-Richardson formula 20, and their respective uses.
 - (c) Work out the coefficient of internal consistency of a test having 36 items, 9.45 as the sum of item variances, and 4.78 as the SD of the total test scores.
11. (a) Give a brief account of different classes of variables involved in an experiment.
 - (b) What are extraneous variables ? Describe different classes of extraneous variables with examples.
 - (c) What are intervening variables.
12. (a) Discuss the relations between the reliability and the validity of a test.
 - (b) What are criterion-related validities. Describe different types of criterion-related validities.
 - (c) What are psychological constructs ? Give an account of construct validity and its types.

13. (a) Use the Spearman-Brown formula to explain how the reliability of a test is related to the length of the test.
- (b) Compare the effects of lengthening of the test length on its validity and reliability.
- (c) A performance test has 40 test items, a reliability coefficient of 0.45 and a validity coefficient of 0.52. What would happen to its reliability and validity if its length is increased to 60 ? Also work out its length required for a validity of 0.60.
14. (a) Mention the basis of content validity and describe what it estimates.
- (b) What are the different ways of working out content validity ?
- (c) Discuss the merits and demerits of the split-half method and of the alternate-forms method of estimating reliability.
15. (a) Describe briefly the working out of the content validity for a test for job performance, and that of the construct validity for an intelligence test.
- (b) Describe the reliability of a psychological test in terms of the variances of the test scores. What is the *SE* of measurement ?
- (c) The reliability coefficient and the *SD* of the total test scores were found to be 0.65 and 8.5 respectively. Compute the true variance and the error variance of the test scores, the *SE* of measurement, and the proportion of the error variance in the total variance.
16. (a) What is a norm ? State briefly its significance in psychological tests.
- (b) Write briefly about nonlinear and linear transformations of test scores with examples.
- (c) Describe the working out, advantages and limitations of the *T* scores.
17. (a) Compare the *C* scale and the stanine scale, mentioning their advantages and limitations.
- (b) Describe the significance of factor analysis in psychological test construction.
- (c) Discuss mentioning the computational formulae how the factor theory theorems can be used in working out validities of tests.

CHAPTER 11

1. Mark the odd item in each of the following series.
- (a) : (i) Scheffe's *F* test, (ii) Gabriel's *SS-STP*, (iii) multiple comparison *t* test, (iv) Bonferroni-modified *t* test.
- (b) : (i) error variance, (ii) between-columns variance, (iii) interaction variance, (iv) between-rows variance.
- (c) : (i) reciprocal transformation, (ii) logarithmic transformation, (iii) square-root transformation, (iv) linear transformation.
- (d) : (i) Kruskal-Wallis *H*, (ii) chi square, (iii) Scheffe's *F*, (iv) Mann-Whitney *U*.
2. Complete each of the following statements by choosing and marking by a tick (✓) mark the correct alternative out of those given below.
- (a) A significant *F* ratio in a one-way model I anova for more than two groups has to be followed by :
- (i) multiple comparison test, (ii) omega square computation, (iii) both, (iv) neither of these.

- (b) A significant F ratio in a one-way Model II anova with two groups has to be followed by :
 (i) multiple comparison t test, (ii) omega square computation, (iii) neither of the two, (iv) both (i) and (ii).
- (c) Where two treatment variables have been used in an experiment, the anova to be used should be :
 (i) two-way model I anova, (ii) one-way model I anova, (iii) one-way model II anova, (iv) two-way model II anova.
- (d) A significant F ratio in a one-way model I anova for only two groups should be followed by :
 (i) Scheffe's F test, (ii) omega square computation, (iii) working out of added variance component, (iv) both (i) and (ii).
- (e) A significant Kruskal-Wallis H in a model I anova for more than two groups has to be followed by :
 (i) Gabriel's $SS-STP$; (ii) Scheffe's F test, (iii) Bonferroni-modified t test, (iv) Mann-Whitney U test.
- (f) The order effect may be eliminated by :
 (i) random sampling, (ii) arc-sine transformation, (iii) random sequence of treatment levels, (iv) none of these.
3. Fill up the blanks in each of the following statements by correct words chosen from those given within the parentheses below.
- (a) The F_i ratio in a two-way anova is the ratio between the _____ variance and the _____ variance.
- (b) Homoscedasticity implies that the groups used in an experiment initially possess _____ of variances, the differences between the latter being due to their _____ errors only.
- (c) In arc-sine transformation, the transformed score is the _____ whose _____ is the square-root of the raw score.
- (d) The F ratio in a one-way anova has the _____ variance as its numerator and the _____ variance as its denominator.
- (e) Where only _____ variables have been applied in the experiment, a _____ model anova has to be used.
- (f) Scheffe's F test is an _____ multiple comparison test while Gabriel's $SS-STP$ is an _____ multiple comparison test.
- (g) The F_c ratio in a two-way model I anova has the _____ variance as the numerator while the F_r ratio in it has the _____ variance as its numerator.
- (a-priori, within-cells, fixed, within-groups, type II, interaction, between-rows, after-design, sine, between-columns; homogeneity, between-groups, angle, sampling, treatment)
4. Match each item of Column 1 with the correct item of Column 2 and put the serial number of the latter in the space after the former.

Column 1

- (a) Scheffe's F test _____
 (b) one-way anova _____
 (c) interaction variance _____
 (d) model II anova _____
 (e) Gabriel's $SS-STP$ _____
 (f) Bonferroni modification _____
 (g) model I anova _____
 (h) Kruskal-Wallis H _____

Column 2

- (i) multiple-comparison t test
 (ii) Mann-Whitney test
 (iii) classification variable
 (iv) a-posteriori test
 (v) single independent variable
 (vi) two-way anova
 (vii) lesser mean square.
 (viii) before-design multiple comparison
 (ix) treatment variable.

5. (a) Discuss the assumptions underlying the one-way anova.

(b) Why is anova preferable to t test ?

(c) Work out one-way anova to find whether or not there is a significant difference between the hourly oxygen consumptions (ml per 100 g bodyweight) of the following sample of parakeets, respectively before and after exposure to a pesticide.

Individual	:	1	2	3	4	5	6	7	8	9	10
Before pesticide:		170	185	160	155	175	168	173	180	165	150
After pesticide :		155	160	130	130	135	148	145	158	148	128

If there be a significant difference, work out the strength of association between oxygen consumption and the levels of pesticide.

6. (a) Explain with examples what you mean by one-way anova, model I anova and model II anova.

(b) Describe how you would work out the variance ratio from the dependent variable scores in a one-way anova and interpret it.

(c) Apply one-way anova to the following pulmonary ventilation data (L/min) of respectively men and women to find if the mean ventilation differs significantly between the sexes. In case of a significant difference, work out the added variance component between the groups.

Men	:	6.55,	7.50,	7.26,	9.00,	8.50,	6.25,	7.30,	8.20,	7.45,	8.25,	8.00.
Women	:	6.10,	6.00,	5.80,	6.35,	6.00,	5.50,	5.75,	6.15,	5.30.		

7. (a) Explain what are *a-priori* and *a-posteriori* multiple comparison tests.

(b) Describe how you would work out the Bonferroni-modified multiple comparison t test and the Scheffe's F test and interpret the computed statistics.

(c) Apply one-way anova to find whether or not there are significant differences between the performance test scores of the following three groups of students after they have practised according to three respective practice schedules.

Group I	:	10,	17,	19,	16,	18,	12,	15,	13,	19,	11.
Group II	:	19,	22,	24,	23,	27,	18,	25,	20,	29,	31.
Group III	:	29,	28,	33,	30,	29,	32,	34,	35,	27.	

If there be significant differences, (i) apply Scheffe's F test to find if the means of Groups II and III differ significantly, and (ii) work out the strength of association of the test scores with the practice schedule.

8. (a) When and how you would work out omega square and added variance component ?

(b) Describe the computation and interpretation of an after-design multiple comparison test.

(c) Following are the amounts of ethereal sulfates (mg) in 24 hours' urine of three groups of humans, kept on low-, medium- and high-protein diets respectively. Use one-way anova to find if dietary protein contents cause significant differences in the urinary ethereal sulfate excretions.

Group I	:	90,	102,	100,	106,	110,	94,	98,	85,	100,	83.
Group II	:	120,	118,	123,	125,	129,	132,	136,	131,	127,	130.
Group III	:	141,	143,	138,	144,	139,	146,	150,	152,	140,	139.

In case of significant differences, (i) apply multiple comparison t test with Bonferroni modification to find if the means of Groups I and II differ significantly, and (ii) work out the strength of association between urinary ethereal sulfates and dietary protein contents.

9. (a) Explain what is meant by a two-way anova without replication.
- (b) Describe, mentioning computational formulae, how you would partition the sum of squares and work out the variance ratios in a two-way anova without replication.
- (c) Apply two-way anova without replication to the following three groups of tracheal ventilation scores (ml/min) of a sample of beetles, exposed to three successive levels of sulfur di-oxide, to find (i) if sulfur di-oxide significantly changes the tracheal ventilation scores, and (ii) if there is a significant added variance component due to random factors between the individual animals.

Individual :	1	2	3	4	5	6	7	8	9	10
Group I :	81.5	85.0	80.7	78.5	80.0	84.5	83.5	79.8	80.4	85.0.
Group II :	70.5	77.2	70.3	67.8	69.2	71.5	71.0	67.3	70.1	77.5.
Group III :	58.6	65.0	59.1	55.7	57.4	59.6	61.2	55.4	58.5	64.0.

10. (a) When do you use two-way anova with replications ?
- (b) Describe how the sum of squares is partitioned and different F ratios are worked out in a two-way model I anova with replications.
- (c) Apply a two-way model I anova to the following fasting blood sugar values (mg/dL) of four groups of subjects, each consisting of 5 individuals and administered one of the four combinations of the levels of cortisol and thyroxine, to find the significance of effects of the treatment variables and of their interaction. Also work out the omega squares for the significant effects.

Fasting blood sugar values		
Levels of thyroxine	Levels of cortisol	
	level 1	level 2
level 1	80	140
	83	152
	85	135
	65	136
	90	145
level 2	120	180
	118	168
	108	176
	112	175
	125	184

11. (a) Write about the assumptions and uses of the Kruskal-Wallis H .
- (b) How do you work out and interpret Kruskal-Wallis anova in case of an experiment with three groups of subjects ?
- (c) Use the Kruskal-Wallis anova to find whether or not there are significant differences between the following heart rates of three groups of workers, performing three respective levels of physical work. If found significant in this test, find whether or not the means of groups I and III differ significantly.

Group I :	80,	90,	85,	80,	84,	102,	95,	82,	100.
Group II :	120,	102,	123,	112,	130,	125,	130,	127.	
Group III :	130,	142,	150,	145,	148,	139,	150,	140.	

Answers*Chapter 1 :*

1. (a) - (iii); (b) - (ii); (c) - (ii); (d) - (iii); (e) - (ii); (f) - (i); (g) - (i).
2. (a) - (ix); (b) - (vii); (c) - (v); (d) - (i); (e) - (ii); (f) - (viii); (g) - (iii); (h) - (vi).
3. (a) relative, treatment; (b) independent, dependent; (c) absolute, classification; (d) homogeneous, stratified; (e) parameter, statistic; (f) probability, choice.
4. (a) - (ii); (b) - (iv); (c) - (i); (d) - (iii).

Chapter 2 :

1. (a) - (iii); (b) - (ii); (c) - (iii); (d) - (ii); (e) - (i); (f) - (iii); (g) - (ii).
2. (a) - (iv); (b) - (i); (c) - (vi); (d) - (v); (e) - (ii).
3. (a) measurement, scattergram; (b) continuous, midpoint; (c) cf, X_u ; (d) many, jaggedness.
4. (a) - (i); (b) - (iv); (c) - (ii); (d) - (iii).

Chapter 3 :

1. (a) - (ii); (b) - (i); (c) - (iii); (d) - (iii); (e) - (ii).
2. (a) median, mean; (b) mode, mean; (c) half, one-tenth; (d) all, 0.04.

Chapter 4 :

1. (a) - (iii); (b) - (i); (c) - (ii); (d) - (iii); (e) - (ii).
2. (a) df, sample-size; (b) relative, absolute; (c) second, first; (d) clumped, relative; (e) squared, no.
3. (a) - (iii); (b) - (ii); (c) - (i); (d) - (iv).

Chapter 5 :

1. (a) - (ii); (b) - (iii); (c) - (i); (d) - (iii); (e) - (i); (f) - (ii).
2. (a) - (ii); (b) - (i); (c) - (iv).
3. (a) statistic, parameter; (b) theoretically, experimentally; (c) parameter, sampling; (d) linearly, standard; (e) nonlinear, linear.

Chapter 6 :

1. (a) - (iii); (b) - (i); (c) - (iii); (d) - (ii); (e) - (ii); (f) - (iii); (g) - (i); (h) - (ii).
2. (a) - (i); (b) - (iv); (c) - (i); (d) - (iv); (e) - (iv).
3. (a) - (v); (b) - (iii); (c) - (vii); (d) - (i); (e) - (iv); (f) - (ii).
4. (a) no, positive; (b) mean, zero; (c) negatively, higher; (d) Gossett, Gauss; (e) less, equals; (f) Poisson, variance; (g) positively, not; (h) half, equals.

Chapter 7 :

1. (a) - (i); (b) - (ii); (c) - (iii); (d) - (ii); (e) - (iii); (f) - (ii).
2. (a) - (iv); (b) - (viii); (c) - (vii); (d) - (iii); (e) - (ix); (f) - (ii); (g) - (v); (h) - (i).
3. (a) rise, rise; (b) retained, acceptance; (c) one-tail, two-tail; (d) acceptance, rejection; (e) accepted, rejection; (f) rise, fall; (g) half, whole.

Chapter 8 :

1. (a) - (ix); (b) - (v); (c) - (i); (d) - (x); (e) - (iii); (f) - (viii); (g) - (vii); (h) - (vi); (i) - (ii).
2. (a) - (iv); (b) - (ii); (c) - (ii); (d) - (iv); (e) - (iv); (f) - (i).
3. (a) *t*, variance; (b) partial, regression; (c) *SE*, estimate; (d) dichotomous, ordinal; (e) negatively, not; (f) rho, tau.
4. (a) - (ii); (b) - (i); (c) - (iv); (d) - (iii).

Chapter 9 :

1. (a) - (iii); (b) - (i); (c) - (iv); (d) - (ii).
2. (a) - (iv); (b) - (iii); (c) - (v); (d) - (i); (e) - (vi).
3. (a) chi-square, *G*; (b) independent, unequal; (c) median, median; (d) independence, contingency; (e) composite, signed; (f) increased, decreased; (g) exceeds, more.
4. (a) - (iv); (b) - (i); (c) - (ii); (d) - (iii); (e) - (iii).

Chapter 10 :

1. (a) systematic, random; (b) total, true; (c) self-correlation, test-whole; (d) reliability, validity; (e) true, common-factor; (f) relevant, independent; (g) equivalent, alternate-forms; (h) intrinsic, reliability.
2. (a) - (iv); (b) - (vi); (c) - (ii); (d) - (vii); (e) - (viii); (f) - (i); (g) - (iii).
3. (a) - (i); (b) - (iii); (c) - (iii); (d) - (ii); (e) - (i).
4. (a) random, systematic; (b) homogeneity, equivalence; (c) decreases, increases; (d) independent, subject-relevant; (e) reliability, validity; (f) concurrent, predictive; (g) identical, high; (h) differing, low.

Chapter 11 :

1. (a) - (ii); (b) - (i); (c) - (iv); (d) - (iii).
2. (a) - (iii); (b) - (iii); (c) - (i); (d) - (ii); (e) - (iv); (f) - (iii).
3. (a) interaction, within-cells; (b) homogeneity, sampling; (c) angle, sine; (d) between-groups, within-groups; (e) treatment, fixed; (f) a-priori, after-design; (g) between-columns, between rows.
4. (a) - (viii); (b) - (v); (c) - (vi); (d) - (iii); (e) - (iv); (f) - (i); (g) - (ix); (h) - (ii).

APPENDIX

Table A. Ordinates at specific x/σ or z scores and areas from the mean to the z scores of the unit normal curve.

$\frac{x}{\sigma}$	Area	Ordinate	$\frac{x}{\sigma}$	Area	Ordinate	$\frac{x}{\sigma}$	Area	Ordinate	$\frac{x}{\sigma}$	Area	Ordinate
.00	.0000	.3989	.55	.2088	.3429	1.10	.3643	.2179	1.65	.4505	.1023
.01	.0040	.3989	.56	.2173	.3410	1.11	.3665	.2155	1.66	.4515	.1006
.02	.0080	.3989	.57	.2157	.3391	1.12	.3686	.2131	1.67	.4525	.0989
.03	.0120	.3988	.58	.2190	.3372	1.13	.3708	.2107	1.68	.4535	.0973
.04	.0160	.3986	.59	.2224	.3352	1.14	.3729	.2083	1.69	.4545	.0957
.05	.0199	.3984	.60	.2257	.3332	1.15	.3749	.2059	1.70	.4554	.0940
.06	.0239	.3982	.61	.2291	.3312	1.16	.3770	.2036	1.71	.4564	.0925
.07	.0279	.3980	.62	.2324	.3292	1.17	.3790	.2012	1.72	.4573	.0909
.08	.0319	.3977	.63	.2357	.3271	1.18	.3810	.1989	1.73	.4582	.0893
.09	.0359	.3973	.64	.2389	.3251	1.19	.3830	.1965	1.74	.4591	.0878
.10	.0398	.3970	.65	.2422	.3230	1.20	.3849	.1942	1.75	.4599	.0863
.11	.0438	.3965	.66	.2454	.3209	1.21	.3869	.1919	1.76	.4608	.0848
.12	.0478	.3961	.67	.2486	.3187	1.22	.3888	.1895	1.77	.4616	.0833
.13	.0517	.3956	.68	.2517	.3166	1.23	.3907	.1872	1.78	.4625	.0818
.14	.0557	.3951	.69	.2549	.3144	1.24	.3925	.1849	1.79	.4633	.0804
.15	.0596	.3945	.70	.2580	.3123	1.25	.3944	.1826	1.80	.4641	.0790
.16	.0636	.3939	.71	.2611	.3101	1.26	.3962	.1804	1.81	.4649	.0775
.17	.0675	.3932	.72	.2642	.3070	1.27	.3980	.1781	1.82	.4656	.0761
.18	.0714	.3925	.73	.2673	.3056	1.28	.3997	.1758	1.83	.4664	.0748
.19	.0753	.3918	.74	.2703	.3034	1.29	.4015	.1736	1.84	.4671	.0734
.20	.0793	.3910	.75	.2734	.3011	1.30	.4032	.1714	1.85	.4678	.0721
.21	.0832	.3902	.76	.2764	.2989	1.31	.4049	.1691	1.86	.4686	.0707
.22	.0871	.3894	.77	.2794	.2966	1.32	.4066	.1669	1.87	.4693	.0694
.23	.0910	.3885	.78	.2823	.2943	1.33	.4082	.1647	1.88	.4699	.0681
.24	.0948	.3876	.79	.2852	.2920	1.34	.4099	.1626	1.89	.4706	.0669
.25	.0987	.3867	.80	.2881	.2897	1.35	.4115	.1604	1.90	.4713	.0656
.26	.1026	.3857	.81	.2910	.2874	1.36	.4131	.1582	1.91	.4719	.0644
.27	.1064	.3847	.82	.2939	.2850	1.37	.4147	.1561	1.92	.4726	.0632
.28	.1103	.3836	.83	.2967	.2827	1.38	.4162	.1539	1.93	.4732	.0620
.29	.1141	.3825	.84	.2995	.2803	1.39	.4177	.1518	1.94	.4738	.0608
.30	.1179	.3814	.85	.3023	.2780	1.40	.4192	.1497	1.95	.4744	.0596
.31	.1217	.3802	.86	.3051	.2756	1.41	.4207	.1476	1.96	.4750	.0584
.32	.1255	.3790	.87	.3078	.2732	1.42	.4222	.1456	1.97	.4756	.0573
.33	.1293	.3778	.88	.3106	.2709	1.43	.4236	.1435	1.98	.4761	.0562
.34	.1331	.3765	.89	.3133	.2685	1.44	.4251	.1415	1.99	.4767	.0551
.35	.1368	.3752	.90	.3159	.2661	1.45	.4265	.1394	2.00	.4772	.0540
.36	.1406	.3739	.91	.3186	.2637	1.46	.4279	.1374	2.01	.4778	.0529
.37	.1443	.3725	.92	.3212	.2613	1.47	.4292	.1354	2.02	.4783	.0519
.38	.1480	.3712	.93	.3238	.2589	1.48	.4306	.1334	2.03	.4788	.0508
.39	.1517	.3697	.94	.3264	.2565	1.49	.4319	.1315	2.04	.4793	.0498
.40	.1554	.3683	.95	.3289	.2541	1.50	.4332	.1295	2.05	.4798	.0488
.41	.1591	.3668	.96	.3315	.2516	1.51	.4345	.1276	2.06	.4803	.0478
.42	.1628	.3653	.97	.3340	.2492	1.52	.4357	.1257	2.07	.4808	.0468
.43	.1664	.3637	.98	.3365	.2468	1.53	.4370	.1238	2.08	.4812	.0459
.44	.1700	.3621	.99	.3389	.2444	1.54	.4382	.1219	2.09	.4817	.0449
.45	.1736	.3605	1.00	.3413	.2420	1.55	.4394	.1200	2.10	.4821	.0440
.46	.1772	.3589	1.01	.3438	.2396	1.56	.4406	.1182	2.11	.4826	.0431
.47	.1808	.3572	1.02	.3461	.2371	1.57	.4418	.1163	2.12	.4830	.0422
.48	.1844	.3555	1.03	.3485	.2347	1.58	.4429	.1145	2.13	.4834	.0413
.49	.1879	.3538	1.04	.3508	.2323	1.59	.4441	.1127	2.14	.4838	.0404
.50	.1915	.3521	1.05	.3531	.2299	1.60	.4452	.1109	2.15	.4842	.0395
.51	.1950	.3503	1.06	.3554	.2275	1.61	.4463	.1092	2.16	.4846	.0387
.52	.1985	.3485	1.07	.3577	.2251	1.62	.4474	.1074	2.17	.4850	.0379
.53	.2019	.3467	1.08	.3599	.2227	1.63	.4484	.1057	2.18	.4854	.0371
.54	.2054	.3448	1.09	.3621	.2203	1.64	.4495	.1040	2.19	.4857	.0363

Table A. Ordinates and areas of the unit normal curve (continued).

$\frac{x}{\sigma}$	Area	Ordinate	$\frac{x}{\sigma}$	Area	Ordinate	$\frac{x}{\sigma}$	Area	Ordinate	$\frac{x}{\sigma}$	Area	Ordinate
2.20	.4861	.0355	2.50	.4938	.0175	2.80	.4974	.0079	3.10	.4990	.0033
2.21	.4864	.0347	2.51	.4940	.0171	2.81	.4975	.0077	3.11	.4991	.0032
2.22	.4868	.0339	2.52	.4941	.0167	2.82	.4976	.0075	3.12	.4991	.0031
2.23	.4871	.0332	2.53	.4943	.0163	2.83	.4977	.0073	3.13	.4991	.0030
2.24	.4875	.0325	2.54	.4945	.0158	2.84	.4977	.0071	3.14	.4992	.0029
2.25	.4878	.0317	2.55	.4946	.0154	2.85	.4978	.0069	3.15	.4992	.0028
2.26	.4881	.0310	2.56	.4948	.0151	2.86	.4979	.0067	3.16	.4992	.0027
2.27	.4884	.0303	2.57	.4949	.0147	2.87	.4979	.0065	3.17	.4992	.0026
2.28	.4887	.0297	2.58	.4951	.0143	2.88	.4980	.0063	3.18	.4993	.0025
2.29	.4890	.0290	2.59	.4952	.0139	2.89	.4981	.0061	3.19	.4993	.0025
2.30	.4893	.0283	2.60	.4953	.0136	2.90	.4981	.0060	3.20	.4993	.0024
2.31	.4896	.0277	2.61	.4955	.0132	2.91	.4982	.0058	3.21	.4993	.0023
2.32	.4898	.0270	2.62	.4956	.0129	2.92	.4982	.0056	3.22	.4994	.0022
2.33	.4901	.0264	2.63	.4957	.0126	2.93	.4983	.0055	3.23	.4994	.0022
2.34	.4904	.0258	2.64	.4959	.0122	2.94	.4984	.0053	3.24	.4994	.0021
2.35	.4906	.0252	2.65	.4960	.0119	2.95	.4984	.0051	3.25	.4994	.0020
2.36	.4909	.0246	2.66	.4961	.0116	2.96	.4985	.0050	3.26	.4994	.0020
2.37	.4911	.0241	2.67	.4962	.0113	2.97	.4985	.0048	3.27	.4995	.0019
2.38	.4913	.0235	2.68	.4963	.0110	2.98	.4986	.0047	3.28	.4995	.0018
2.39	.4916	.0229	2.69	.4964	.0107	2.99	.4986	.0046	3.29	.4995	.0018
2.40	.4918	.0224	2.70	.4965	.0104	3.00	.4987	.0044	3.30	.4995	.0017
2.41	.4920	.0219	2.71	.4966	.0101	3.01	.4987	.0043	3.40	.4997	.0012
2.42	.4922	.0213	2.72	.4967	.0099	3.02	.4987	.0042	3.50	.4998	.0009
2.43	.4925	.0208	2.73	.4968	.0096	3.03	.4988	.0040	3.60	.4998	.0006
2.44	.4927	.0203	2.74	.4969	.0093	3.04	.4988	.0039	3.70	.4999	.0004
2.45	.4929	.0198	2.75	.4970	.0091	3.05	.4989	.0038	3.80	.49993	.0003
2.46	.4931	.0194	2.76	.4971	.0088	3.06	.4989	.0037	3.90	.49995	.0002
2.47	.4932	.0189	2.77	.4972	.0086	3.07	.4989	.0036	4.00	.49997	.0001
2.48	.4934	.0184	2.78	.4973	.0084	3.08	.4990	.0035			
2.49	.4936	.0180	2.79	.4974	.0081	3.09	.4990	.0034			

Table reproduced by permission of McGraw-Hill Book Company, New York, from G.A. Ferguson, *Statistical Analysis in Psychology and Education*, 3rd ed., 1971, and originally from J. E. Wert, *Educational Statistics*, both published by McGraw-Hill Book Company.

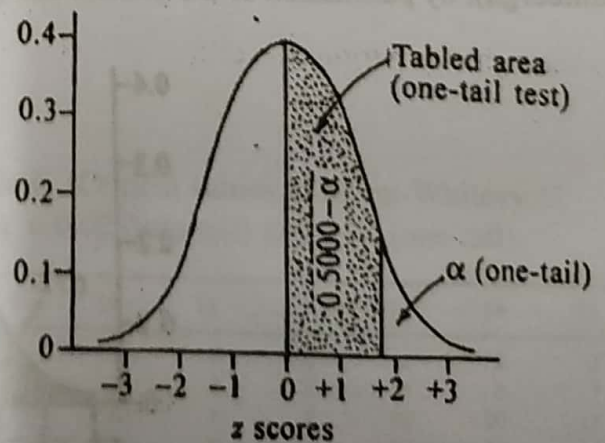
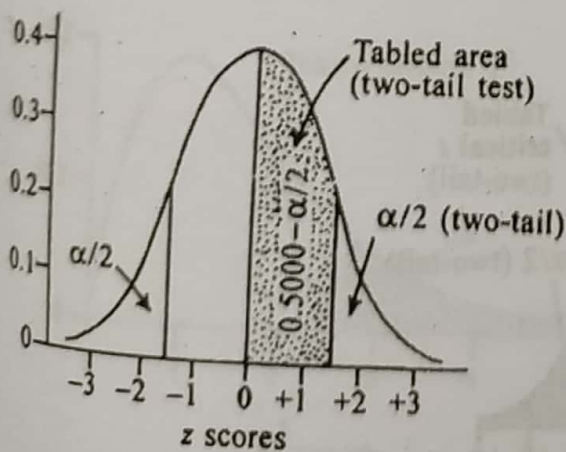


Table B. Critical values of t .

Levels of significance for two-tail test

df	.10	.05	.02	.01	.001	df	.10	.05	.02	.01	.001
1	6.314	12.706	31.821	63.657	636.619	34	1.6991	2.032	2.441	2.728	3.601
2	2.920	4.303	6.965	9.925	31.598	35	1.6990	2.030	2.438	2.724	3.591
3	2.353	3.182	4.541	5.841	12.941	36	1.6888	2.028	2.434	2.719	3.582
4	2.132	2.776	3.747	4.604	8.610	37	1.6887	2.026	2.431	2.715	3.574
5	2.015	2.571	3.365	4.032	6.859	38	1.6886	2.024	2.429	2.712	3.566
6	1.943	2.447	3.143	3.707	5.959	39	1.6885	2.023	2.426	2.708	3.558
7	1.895	2.365	2.998	3.499	5.405	40	1.684	2.021	2.423	2.704	3.551
8	1.860	2.306	2.896	3.355	5.041	41	1.683	2.020	2.421	2.701	3.544
9	1.833	2.262	2.821	3.250	4.781	42	1.682	2.018	2.418	2.698	3.538
10	1.812	2.228	2.764	3.169	4.587	43	1.681	2.017	2.416	2.695	3.532
11	1.796	2.201	2.718	3.106	4.437	44	1.680	2.015	2.414	2.692	3.526
12	1.782	2.179	2.681	3.055	4.318	45	1.679	2.014	2.412	2.690	3.520
13	1.771	2.160	2.650	3.012	4.221	46	1.679	2.013	2.410	2.687	3.515
14	1.761	2.145	2.624	2.977	4.140	47	1.678	2.012	2.408	2.685	3.510
15	1.753	2.131	2.602	2.947	4.073	48	1.677	2.011	2.407	2.682	3.505
16	1.746	2.120	2.583	2.921	4.015	49	1.677	2.010	2.405	2.680	3.500
17	1.740	2.110	2.567	2.898	3.965	50	1.676	2.009	2.403	2.678	3.496
18	1.734	2.101	2.552	2.878	3.922	51	1.675	2.008	2.402	2.676	3.492
19	1.729	2.093	2.539	2.861	3.883	52	1.675	2.007	2.400	2.674	3.488
20	1.725	2.086	2.528	2.845	3.850	53	1.674	2.006	2.399	2.672	3.484
21	1.721	2.080	2.518	2.831	3.819	54	1.674	2.005	2.397	2.670	3.480
22	1.717	2.074	2.508	2.819	3.792	55	1.673	2.004	2.396	2.668	3.476
23	1.714	2.069	2.500	2.807	3.767	56	1.673	2.003	2.395	2.667	3.473
24	1.711	2.064	2.492	2.797	3.745	57	1.672	2.002	2.394	2.665	3.470
25	1.708	2.060	2.485	2.787	3.725	58	1.672	2.002	2.392	2.663	3.466
26	1.706	2.056	2.479	2.779	3.707	59	1.671	2.001	2.391	2.662	3.463
27	1.703	2.052	2.473	2.771	3.690	60	1.671	2.000	2.390	2.660	3.460
28	1.701	2.048	2.467	2.763	3.674	70	1.667	1.994	2.381	2.648	3.435
29	1.699	2.045	2.462	2.756	3.659	80	1.664	1.990	2.374	2.639	3.416
30	1.697	2.042	2.457	2.750	3.646	90	1.662	1.987	2.368	2.632	3.402
31	1.696	2.040	2.453	2.744	3.633	100	1.660	1.984	2.364	2.626	3.390
32	1.6994	2.037	2.449	2.738	3.622	120	1.658	1.980	2.358	2.617	3.373
33	1.6992	2.035	2.445	2.733	3.611	∞	1.645	1.960	2.326	2.576	3.291
	.05	.025	.01	.005	.0005		.05	.025	.01	.005	.0005

Levels of significance for one-tail test

Table taken from Table III of R.A. Fisher and F.Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, published by Longman Group Ltd., London (previously published by Oliver & Boyd, Edinburgh), by permission of the authors and publishers.

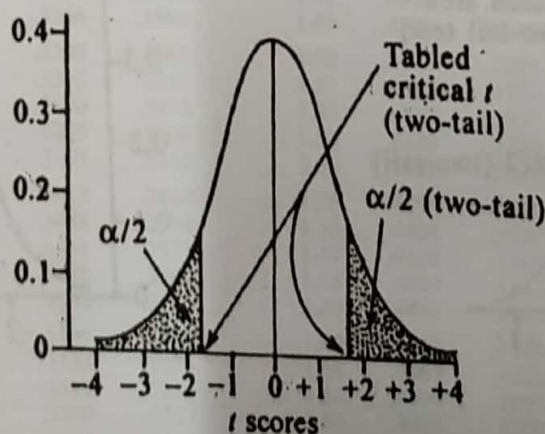
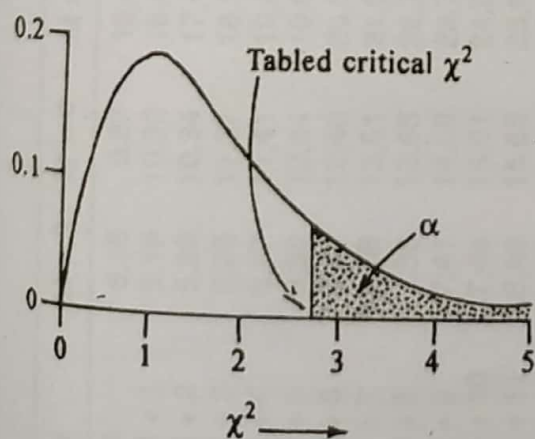


Table C. Critical values of chi square. Figures at the top of the table indicate levels of significance (α).

df	.50	.30	.20	.10	.05	.02	.01	.001
1	0.46	1.07	1.64	2.71	3.84	5.41	6.64	10.83
2	1.39	2.41	3.22	4.60	5.99	7.82	9.21	13.82
3	2.37	3.66	4.64	6.25	7.82	9.84	11.34	16.27
4	3.36	4.88	5.99	7.78	9.49	11.67	13.28	18.46
5	4.35	6.06	7.29	9.24	11.07	13.39	15.09	20.52
6	5.35	7.23	8.56	10.64	12.59	15.03	16.81	22.46
7	6.35	8.38	9.80	12.02	14.07	16.62	18.48	24.32
8	7.34	9.52	11.03	13.36	15.51	18.17	20.09	26.12
9	8.34	10.66	12.24	14.68	16.92	19.68	21.67	27.88
10	9.34	11.78	13.44	15.99	18.31	21.16	23.21	29.59
11	10.34	12.90	14.63	17.28	19.68	22.62	24.72	31.26
12	11.34	14.01	15.81	18.55	21.03	24.05	26.22	32.91
13	12.34	15.12	16.98	19.81	22.36	25.47	27.69	34.53
14	13.34	16.22	18.15	21.06	23.68	26.87	29.14	36.12
15	14.34	17.32	19.31	22.81	25.00	28.26	30.58	37.70
16	15.34	18.42	20.46	23.54	26.30	29.63	32.00	39.29
17	16.34	19.51	21.62	24.77	27.59	31.00	33.41	40.75
18	17.34	20.60	22.76	25.99	28.87	32.35	34.80	42.31
19	18.34	21.69	23.90	27.20	30.14	33.69	36.19	43.82
20	19.34	22.78	25.04	28.41	31.41	35.02	37.57	45.82
21	20.34	23.86	26.17	29.62	32.67	36.34	38.93	46.80
22	21.34	24.94	27.30	30.81	33.92	37.66	40.29	48.27
23	22.34	26.02	28.43	32.01	35.17	38.97	41.64	49.73
24	23.34	27.10	29.55	33.20	36.42	40.27	42.98	51.18
25	24.34	28.17	30.68	34.38	37.65	41.57	44.31	52.62
26	25.34	29.25	31.80	35.56	38.88	42.86	45.64	54.05
27	26.34	30.32	32.91	36.74	40.11	44.14	46.96	55.48
28	27.34	31.39	34.03	37.92	41.34	45.42	48.28	56.89
29	28.34	32.46	35.14	39.09	42.56	46.69	49.59	58.30
30	29.34	33.53	36.25	40.26	43.77	47.96	50.89	59.70

Table taken from Table IV of R.A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, published by Longman Group Ltd., London (previously published by Oliver & Boyd, Edinburgh), by permission of the authors and publishers.

**Table D.** Critical values of Spearman's rho.

Levels of significance for two-tail test			
n	.10	.02	.01
5	.900	1.000	
6	.829	.943	1.000
7	.714	.893	.929
8	.643	.833	.881
9	.600	.783	.833
10	.564	.746	.794
12	.506	.712	.777
14	.456	.645	.715
16	.425	.601	.665
18	.399	.564	.625
20	.377	.534	.591
22	.359	.508	.562
24	.343	.485	.537
26	.329	.465	.515
28	.317	.448	.496
30	.306	.432	.478
	.05	.01	.005

Levels of significance for one-tail test

Reprinted by permission of McGraw-Hill Book Company, New York, from J. P. Guilford and B. Fruchter, *Fundamental Statistics in Psychology and Education*, 6th ed., 1978, and originally from W.J. Dixon and F.J. Massey, Jr., *Introduction to Statistical Analysis*, 1951, both published by McGraw-Hill Book Company.

Table E. Critical values of Mann-Whitney U . $\alpha = 0.02$ (two-tail) and 0.01 (one-tail).

n	9	10	11	12	13	14	15
3	1	1	1	2	2	2	3
4	3	3	4	5	5	6	7
5	5	6	7	8	9	10	11
6	7	8	9	11	12	13	15
7	9	11	12	14	16	17	19
8	11	13	15	17	20	22	24

Table F. Critical T_0 scores for Wilcoxon's signed rank test.
 n = size of each sample or group. α = level of significance.

n	α for two-tail tests				n	α for two-tail tests			
	.10	.05	.02	.01		.10	.05	.02	.01
5	1				28	130	117	102	92
6	2	1			29	141	127	111	100
7	4	2	0		30	152	137	120	109
8	6	4	2	0	31	163	148	130	118
9	8	6	3	2	32	175	159	141	128
10	11	8	5	3	33	188	171	151	138
11	14	11	7	5	34	201	183	162	149
12	17	14	10	7	35	214	195	174	160
13	21	17	13	10	36	228	208	186	171
14	26	21	16	13	37	242	222	198	183
15	30	25	20	16	38	256	235	211	195
16	36	30	24	19	39	271	250	224	208
17	41	35	28	23	40	287	264	238	221
18	47	40	33	28	41	303	279	252	234
19	54	46	38	32	42	319	295	267	248
20	60	52	43	37	43	336	311	281	262
21	68	59	49	43	44	353	327	297	277
22	75	66	56	49	45	371	344	313	292
23	83	73	62	55	46	389	361	329	307
24	92	81	69	61	47	408	379	345	323
25	101	90	77	68	48	427	397	362	339
26	110	98	85	76	49	446	415	380	356
27	120	107	93	84	50	466	434	398	373

Table taken from Frank Wilcoxon and Roberta A. Wilcox, *Some Rapid Approximate Statistical Procedure*, revised ed., 1964, Lederle Laboratories, New York, by the kind permission of American Cyanamid Company.

Table G. Critical upper and lower T values (T_u and T_l) for Wilcoxon's rank sum test.
 N, M = sample sizes. $\alpha = 0.005$ (one-tail) and 0.01 (two-tail).

N	$M = 3$	$M = 4$	$M = 5$	$M = 6$	$M = 7$	$M = 8$	$M = 9$	$M = 10$	$M = 11$	$M = 12$	$M = 13$	$M = 14$
$N = M$	5,16	9,27	15,40	23,55	33,72	44,92	57,114	71,139	88,165	106,194	126,225	148,258
$N = M + 1$	5,19	10,30	16,44	24,60	34,78	46,98	59,121	74,146	91,173	109,203	130,234	152,268
$N = M + 2$	5,22	10,34	17,48	25,65	36,83	7,105	61,128	76,154	94,181	113,211	133,244	156,278
$N = M + 3$	5,25	11,37	18,52	27,69	37,89	49,111	63,135	79,161	97,189	116,220	137,253	160,288
$N = M + 4$	6,27	11,41	19,56	28,74	39,94	51,117	65,142	82,168	100,197	119,229	141,262	164,298
$N = M + 5$	6,30	12,44	19,61	29,79	40,100	53,123	68,148	84,176	102,206	123,237	144,272	168,308
$N = M + 6$	6,33	12,48	20,65	30,84	42,105	55,129	70,155	87,183	105,214	126,246	148,281	172,318
$N = M + 7$	6,36	13,51	21,69	31,89	43,111	57,135	72,162	89,191	108,222	129,255	152,290	176,328
$N = M + 8$	7,38	13,55	22,73	32,94	45,116	59,141	74,169	92,198	111,230	133,263	156,299	180,338
$N = M + 9$	7,41	14,58	23,77	34,98	46,122	61,147	77,175	95,205	114,238	136,272	159,309	185,347
$N = M + 10$	7,44	15,61	24,81	35,103	48,127	62,154	79,182	97,213	117,246	139,281	163,318	189,357
$N = M + 11$	8,46	15,65	25,85	36,108	49,133	64,160	81,139	100,220	120,254	143,289	167,327	193,367
$N = M + 12$	8,49	16,68	26,89	37,113	51,138	66,166	83,196	103,227	123,262	146,298	171,336	197,377
$N = M + 13$	8,52	16,72	26,94	38,118	52,144	68,172	86,202	105,235	126,270	150,306	175,345	201,387
$N = M + 14$	9,54	17,75	27,98	40,122	54,149	70,178	88,209	108,242	129,278	153,315	178,355	205,397
$N = M + 15$	9,57	17,79	28,102	41,127	55,155	72,184	90,216	110,250	132,286	156,324	182,364	210,406
$N = M + 16$	9,60	18,82	29,106	42,132	57,160	74,190	93,222	113,257	136,293	160,332	186,373	214,416
$N = M + 17$	9,63	19,85	30,110	43,137	59,165	76,196	95,229	116,264	139,301	163,341	190,382	218,426
$N = M + 18$	10,65	19,89	31,114	45,141	60,171	78,202	97,236	118,272	142,309	167,349	194,391	222,436
$N = M + 19$	10,68	20,92	32,118	46,146	62,176	80,208	99,243	121,279	145,317	170,358	197,401	226,446
$N = M + 20$	10,71	20,96	33,122	47,151	63,182	82,214	102,249	124,286	148,325	173,367	201,410	231,455
$N = M + 21$	11,73	21,99	33,127	48,156	65,187	83,221	104,256	126,294	151,333	177,375	205,419	235,465
$N = M + 22$	11,76	21,103	34,131	49,161	66,193	85,227	106,263	129,301	154,341	180,384	209,428	239,475
$N = M + 23$	11,79	22,106	35,135	51,165	68,198	87,233	109,269	132,308	157,349	184,392	213,437	243,485
$N = M + 24$	12,81	23,109	36,139	52,170	70,203	89,239	111,276	134,316	160,357	187,401	216,447	247,495
$N = M + 25$	12,84	23,113	37,143	53,175	71,209	91,245	113,283	137,323	163,365	191,409	220,456	252,504

Reprinted from Frank Wilcoxon and Roberta A. Wilcox, *Some Rapid Approximate Statistical Procedures*, revised ed., 1964, Lederle Laboratories, New York, by the kind permission of American Cyanamid Company.

Table G. Critical upper and lower T values for Wilcoxon's rank sum test (continued).
 $\alpha = 0.005$ (one-tail) and 0.01 (two-tail).

N	M = 15	M = 16	M = 17	M = 18	M = 19	M = 20	M = 21	M = 22	M = 23	M = 24	M = 25
N = M	171,294	196,332	223,372	252,414	283,458	316,504	350,553	386,604	424,657	464,712	506,769
N = M + 1	176,304	201,343	229,383	258,426	289,471	322,518	357,567	393,619	432,672	472,728	514,786
N = M + 2	180,315	206,354	234,395	264,438	295,484	329,531	364,581	401,633	440,687	480,744	523,802
N = M + 3	184,326	211,365	239,407	269,451	301,497	335,545	371,595	408,648	447,703	489,759	531,819
N = M + 4	189,336	216,376	245,418	275,463	307,510	342,558	378,609	415,663	455,718	497,775	540,835
N = M + 5	194,346	221,387	250,430	281,475	314,522	348,572	385,623	423,677	463,733	505,791	549,851
N = M + 6	198,357	226,398	255,442	287,487	320,535	355,585	392,637	430,692	471,748	513,807	557,868
N = M + 7	203,367	231,409	261,453	292,500	326,548	361,599	399,651	438,706	479,763	521,823	566,884
N = M + 8	207,378	236,420	266,465	298,512	332,561	368,612	405,666	445,721	486,779	530,838	575,900
N = M + 9	212,388	241,431	271,477	304,524	338,574	374,626	412,680	452,736	494,794	538,854	583,917
N = M + 10	216,399	245,443	277,488	310,536	344,587	381,639	419,694	460,750	502,809	546,870	592,933
N = M + 11	221,409	250,454	282,500	315,549	351,599	388,652	426,708	467,765	510,824	554,886	601,949
N = M + 12	225,420	255,465	287,512	321,561	357,612	394,666	433,722	475,779	518,839	563,901	609,966
N = M + 13	230,430	260,476	293,523	327,573	363,625	401,679	440,736	482,794	526,854	571,917	618,982
N = M + 14	235,440	265,487	298,535	333,585	369,638	407,693	447,750	490,808	533,870	579,933	627,998
N = M + 15	239,451	270,498	303,547	338,598	375,651	414,706	454,764	497,823	541,885	587,949	635,1015
N = M + 16	244,461	275,509	309,558	344,610	381,664	421,719	462,777	504,838	549,900	596,964	644,1031
N = M + 17	248,472	280,520	314,570	350,622	388,676	427,733	469,791	512,852	565,930	612,996	661,1064
N = M + 18	253,482	285,531	320,581	356,634	394,689	434,746	476,805	519,867	573,945	620,1012	670,1080
N = M + 19	257,493	290,542	325,593	362,646	400,702	440,760	483,819	527,881	580,961	629,1027	679,1096
N = M + 20	262,503	295,553	330,605	367,659	406,715	447,773	490,833	534,896	588,976	637,1043	687,1113
N = M + 21	267,513	300,564	336,616	373,671	413,727	454,786	497,847	542,910	596,991	645,1059	696,1129
N = M + 22	271,524	305,575	341,628	379,683	419,740	460,800	504,861	549,925	604,1006	654,1074	705,1145
N = M + 23	276,534	310,586	347,639	385,695	425,753	467,813	511,875	556,940	612,1021	662,1090	714,1161
N = M + 24	280,545	315,597	352,651	391,707	431,766	474,826	518,889	564,954	620,1036	670,1106	722,1178
N = M + 25	285,555	320,608	357,663	397,719	438,778	480,840	525,903	571,969			

Table G. Critical upper and lower T values for Wilcoxon's rank sum test (continued).
 $\alpha = 0.025$ (one-tail) and 0.05 (two-tail).

N	$M = 3$	$M = 4$	$M = 5$	$M = 6$	$M = 7$	$M = 8$	$M = 9$	$M = 10$	$M = 11$	$M = 12$	$M = 13$	$M = 14$
$N = M$	5,16	11,25	18,37	26,52	37,68	49,87	63,108	79,131	96,157	116,184	137,214	160,246
$N = M + 1$	6,18	12,28	19,41	28,56	39,73	51,93	66,114	82,138	100,164	120,192	141,223	165,255
$N = M + 2$	6,21	12,32	20,45	29,61	41,78	54,98	68,121	85,145	103,172	124,200	146,231	170,264
$N = M + 3$	7,23	13,35	21,49	31,65	43,83	56,104	71,127	88,152	107,179	128,208	150,240	174,274
$N = M + 4$	7,26	14,38	22,53	32,70	45,88	58,110	74,133	91,159	110,187	131,217	154,249	179,283
$N = M + 5$	8,28	15,41	24,56	34,74	46,94	61,115	77,139	94,166	114,194	135,225	159,257	184,292
$N = M + 6$	8,31	16,44	25,60	36,78	48,99	63,121	79,146	97,173	118,201	139,233	163,266	189,301
$N = M + 7$	9,33	17,47	26,64	37,83	50,104	65,127	82,152	101,179	121,209	143,241	168,274	194,310
$N = M + 8$	10,35	17,51	27,68	39,87	52,109	68,132	85,158	104,186	125,216	147,249	172,283	198,320
$N = M + 9$	10,38	18,54	29,71	41,91	54,114	70,138	88,164	107,193	128,224	151,257	176,292	203,329
$N = M + 10$	11,40	19,57	30,75	42,96	56,119	72,144	90,171	110,200	132,231	155,265	181,300	208,338
$N = M + 11$	11,43	20,60	31,79	44,100	58,124	75,149	93,177	113,207	135,239	159,273	185,309	213,347
$N = M + 12$	12,45	21,63	32,83	45,105	60,129	77,155	96,183	117,213	139,246	163,281	190,317	218,356
$N = M + 13$	12,48	22,66	33,87	47,109	62,134	80,160	99,189	120,220	143,253	167,289	194,326	222,366
$N = M + 14$	13,50	23,69	35,90	49,113	64,139	82,166	101,196	123,227	146,261	171,297	198,335	227,375
$N = M + 15$	13,53	24,72	36,94	50,118	66,144	84,172	104,202	126,234	150,268	175,305	203,343	232,384
$N = M + 16$	14,55	24,76	37,98	52,122	68,149	87,177	107,208	129,241	153,276	179,313	207,352	237,393
$N = M + 17$	14,58	25,79	38,102	53,127	70,154	89,183	110,214	132,248	157,283	183,321	212,360	242,402
$N = M + 18$	15,60	26,82	40,105	55,131	72,159	92,188	113,220	136,254	161,290	187,329	216,369	247,411
$N = M + 19$	15,63	27,85	41,109	57,135	74,164	94,194	115,227	139,261	164,298	191,337	221,377	252,420
$N = M + 20$	16,65	28,88	42,113	58,140	76,169	96,200	118,233	142,268	168,305	195,345	225,386	256,430
$N = M + 21$	16,68	29,91	43,117	60,144	78,174	99,205	121,239	145,275	171,313	199,353	229,395	261,439
$N = M + 22$	17,70	30,94	45,120	61,149	80,179	101,211	124,245	148,282	175,320	203,361	234,403	266,448
$N = M + 23$	17,73	31,97	46,124	63,153	82,184	103,217	127,251	152,288	179,327	207,369	238,412	271,457
$N = M + 24$	18,75	31,101	47,128	65,157	84,189	106,222	129,258	155,295	182,335	211,377	243,420	276,466
$N = M + 25$	18,78	32,104	48,132	66,162	86,194	108,228	132,264	158,302	186,342	216,384	247,429	281,475

Table G. Critical upper and lower T values for Wilcoxon's rank sum test (continued).
 $\alpha = 0.025$ (one-tail) and 0.05 (two-tail).

N	$M = 15$	$M = 16$	$M = 17$	$M = 18$	$M = 19$	$M = 20$	$M = 21$	$M = 22$	$M = 23$	$M = 24$	$M = 25$
$N = M$	185,280	212,316	240,355	271,395	303,438	337,483	373,530	411,579	451,630	483,683	536,739
$N = M + 1$	190,290	217,327	246,366	277,407	310,450	345,495	381,543	419,593	460,644	502,698	546,754
$N = M + 2$	195,300	223,337	252,377	284,418	317,462	352,508	389,556	428,606	468,659	511,713	555,770
$N = M + 3$	201,309	229,347	258,388	290,430	324,474	359,521	397,569	436,620	477,673	520,728	565,785
$N = M + 4$	206,319	234,358	264,399	297,441	331,486	367,533	404,583	444,634	486,687	529,743	574,801
$N = M + 5$	211,329	240,368	271,409	303,453	338,498	374,546	412,596	452,648	494,702	538,758	584,816
$N = M + 6$	216,339	245,379	277,420	310,464	345,510	381,559	420,609	460,662	503,716	547,773	593,832
$N = M + 7$	221,349	251,389	283,431	316,476	351,523	389,571	428,622	469,675	512,730	556,788	603,847
$N = M + 8$	227,358	257,399	289,442	323,487	358,535	396,584	436,635	477,689	520,745	565,803	612,863
$N = M + 9$	232,368	262,410	295,453	329,499	365,547	403,597	443,649	485,703	529,759	575,817	622,878
$N = M + 10$	237,378	268,420	301,464	336,510	372,559	411,609	451,662	493,717	538,773	584,832	632,893
$N = M + 11$	242,388	274,430	307,475	342,522	379,571	418,622	459,675	502,730	546,788	593,847	641,909
$N = M + 12$	248,397	279,441	313,486	349,533	386,583	426,634	467,688	510,744	555,802	602,862	651,924
$N = M + 13$	253,407	285,451	319,497	355,545	393,595	433,647	475,701	518,758	564,816	611,877	660,940
$N = M + 14$	258,417	291,461	325,508	362,556	400,607	440,660	482,715	526,772	572,831	620,892	670,955
$N = M + 15$	263,427	296,472	331,519	368,568	407,619	448,672	490,728	535,785	581,845	629,907	679,971
$N = M + 16$	269,436	302,482	338,529	375,579	414,631	455,685	498,741	543,799	590,859	638,922	689,986
$N = M + 17$	274,446	308,492	344,540	381,591	421,643	463,697	506,754	551,813	599,873	648,936	699,1001
$N = M + 18$	279,456	314,502	350,551	388,602	428,655	470,710	514,767	560,826	607,888	657,951	708,1017
$N = M + 19$	284,466	319,513	356,562	395,613	435,667	477,723	522,780	568,840	616,902	666,966	718,1032
$N = M + 20$	290,475	325,523	362,573	401,625	442,679	485,735	530,793	576,854	625,916	675,981	727,1048
$N = M + 21$	295,485	331,533	368,584	408,636	449,691	492,748	537,807	584,868	633,931	684,996	737,1063
$N = M + 22$	300,495	336,544	374,595	414,648	456,703	500,760	545,820	593,881	642,945	693,1011	747,1063
$N = M + 23$	306,504	342,554	380,606	421,659	463,715	507,773	553,833	601,895	651,959	703,1025	756,1094
$N = M + 24$	311,514	348,564	387,616	427,671	470,727	515,785	561,846	609,909	660,973	712,1040	766,1109
$N = M + 25$	316,524	353,575	393,627	434,682	477,739	522,798	569,859	618,922	668,988	721,1055	775,1125

Table H. Critical values of F . $\alpha = 0.05$ (roman type) and 0.01 (bold-face type).

Degrees of freedom for lesser mean square	Degrees of freedom for greater mean square																												
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞					
1	161 4052	200 4999	216 5402	225 5635	230 5764	234 5859	237 5928	239 5981	241 6022	242 6058	243 6082	244 6106	245 6142	246 6169	248 6208	249 6234	250 6258	251 6286	252 6302	253 6323	253 6334	254 6352	254 6361	254 6366					
2	18.51 98.49	19.00 99.01	19.16 99.17	19.25 99.25	19.30 99.30	19.33 99.33	19.36 99.34	19.37 99.36	19.38 99.38	19.39 99.40	19.40 99.41	19.41 99.42	19.42 99.43	19.43 99.44	19.44 99.45	19.45 99.46	19.46 99.47	19.47 99.48	19.47 99.48	19.48 99.49	19.49 99.49	19.49 99.49	19.50 99.50	19.50 99.50					
3	10.13 34.12	9.55 30.81	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.88 27.67	8.84 27.49	8.81 27.34	8.78 27.23	8.76 27.13	8.74 27.05	8.71 26.92	8.69 26.83	8.66 26.69	8.64 26.60	8.62 26.50	8.60 26.41	8.58 26.35	8.57 26.27	8.56 26.23	8.54 26.18	8.54 26.14	8.53 26.12					
4	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.26 15.52	6.16 15.21	6.09 14.98	6.04 14.80	6.00 14.66	5.96 14.54	5.93 14.45	5.91 14.37	5.87 14.24	5.84 14.15	5.80 14.02	5.77 13.93	5.74 13.83	5.71 13.74	5.70 13.69	5.68 13.61	5.66 13.57	5.65 13.52	5.64 13.48	5.63 13.46					
5	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.29	5.05 10.97	4.95 10.67	4.88 10.45	4.82 10.27	4.78 10.15	4.74 10.05	4.68 9.96	4.64 9.82	4.60 9.77	4.56 9.68	4.53 9.55	4.50 9.47	4.46 9.38	4.44 9.29	4.44 9.24	4.42 9.17	4.40 9.13	4.38 9.07	4.37 9.04	4.36 9.02					
6	5.99 13.74	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.93	4.06 7.87	4.03 7.79	4.00 7.72	3.96 7.60	3.92 7.52	3.87 7.39	3.84 7.31	3.81 7.23	3.77 7.14	3.75 7.09	3.72 7.02	3.71 6.99	3.69 6.94	3.68 6.90	3.67 6.88					
7	5.59 12.25	4.74 9.58	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 7.00	3.73 6.84	3.68 6.71	3.63 6.62	3.60 6.54	3.57 6.47	3.52 6.35	3.49 6.27	3.44 6.15	3.41 6.07	3.38 5.98	3.34 5.90	3.32 5.85	3.29 5.78	3.28 5.75	3.25 5.70	3.24 5.67	3.23 5.65					
8	5.32 11.26	4.46 8.65	4.07 7.89	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.19	3.44 6.05	3.39 5.91	3.34 5.82	3.31 5.74	3.28 5.67	3.23 5.56	3.20 5.48	3.15 5.36	3.12 5.28	3.08 5.20	3.05 5.11	3.03 5.06	3.00 5.00	2.98 4.96	2.96 4.91	2.94 4.88	2.93 4.84					
9	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.62	3.23 5.47	3.18 5.35	3.13 5.26	3.10 5.18	3.07 5.11	3.02 5.00	2.98 4.92	2.93 4.80	2.90 4.73	2.86 4.64	2.82 4.56	2.80 4.51	2.77 4.45	2.76 4.41	2.73 4.36	2.72 4.33	2.71 4.31					
10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.21	3.07 5.06	3.02 4.95	2.97 4.85	2.94 4.78	2.91 4.71	2.86 4.60	2.82 4.52	2.77 4.41	2.74 4.33	2.70 4.25	2.67 4.17	2.64 4.12	2.61 4.05	2.59 4.01	2.56 3.96	2.55 3.93	2.54 3.91					
11	4.84 9.65	3.98 7.20	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.88	2.95 4.74	2.90 4.63	2.86 4.54	2.82 4.46	2.79 4.40	2.74 4.29	2.70 4.21	2.65 4.10	2.61 4.02	2.57 3.94	2.53 3.86	2.50 3.80	2.47 3.74	2.45 3.70	2.42 3.66	2.41 3.62	2.40 3.60					
12	4.75 9.33	3.88 6.93	3.49 5.95	3.26 5.41	3.11 5.06	3.00 4.82	2.92 4.65	2.85 4.50	2.80 4.39	2.76 4.30	2.72 4.22	2.69 4.16	2.64 4.05	2.60 3.98	2.54 3.86	2.50 3.78	2.46 3.70	2.42 3.61	2.40 3.56	2.36 3.48	2.35 3.46	2.32 3.41	2.31 3.38	2.30 3.24					
13	4.67 9.07	3.80 6.70	3.41 5.74	3.18 5.20	3.02 4.86	2.92 4.62	2.84 4.44	2.77 4.30	2.72 4.19	2.67 4.10	2.63 4.02	2.60 3.96	2.55 3.85	2.51 3.78	2.46 3.67	2.42 3.59	2.38 3.51	2.34 3.42	2.32 3.37	2.28 3.36	2.26 3.27	2.24 3.21	2.22 3.18	2.21 3.16					
14	4.60 8.86	3.74 6.51	3.34 5.56	3.11 5.03	2.96 4.69	2.85 4.46	2.77 4.28	2.70 4.14	2.65 4.03	2.60 3.94	2.56 3.86	2.53 3.80	2.48 3.70	2.44 3.64	2.39 3.51	2.35 3.43	2.31 3.34	2.27 3.26	2.24 3.21	2.21 3.14	2.19 3.11	2.16 3.06	2.14 3.02	2.13 3.00					
15	4.54 8.68	3.68 6.36	3.29 5.42	3.06 4.89	2.90 4.56	2.79 4.32	2.70 4.14	2.64 4.00	2.59 3.89	2.55 3.80	2.51 3.73	2.48 3.67	2.43 3.56	2.39 3.48	2.33 3.36	2.29 3.29	2.25 3.20	2.21 3.12	2.18 3.07	2.15 3.00	2.12 2.97	2.10 2.92	2.08 2.89	2.07 2.87					
16	4.49 8.53	3.63 6.23	3.24 5.29	3.01 4.77	2.85 4.44	2.74 4.20	2.66 4.03	2.59 3.89	2.54 3.78	2.49 3.69	2.45 3.61	2.42 3.55	2.37 3.45	2.33 3.37	2.28 3.25	2.24 3.15	2.20 3.10	2.16 3.01	2.13 2.96	2.09 2.89	2.07 2.84	2.04 2.80	2.02 2.77	2.01 2.75					
17	4.45 8.40	3.59 6.11	3.20 5.18	2.96 4.67	2.81 4.34	2.70 4.10	2.62 3.93	2.55 3.79	2.50 3.68	2.45 3.59	2.41 3.52	2.38 3.45	2.33 3.35	2.29 3.27	2.23 3.16	2.19 3.08	2.15 3.00	2.11 2.92	2.08 2.86	2.04 2.79	2.02 2.76	1.99 2.70	1.97 2.67	1.96 2.65					
18	4.41 8.28	3.55 6.01	3.16 5.09	2.93 4.58	2.77 4.25	2.66 4.01	2.58 3.85	2.51 3.71	2.46 3.60	2.41 3.51	2.37 3.44	2.34 3.37	2.29 3.27	2.25 3.19	2.19 3.07	2.15 3.00	2.11 2.91	2.07 2.83	2.04 2.78	2.00 2.71	1.98 2.68	1.95 2.62	1.93 2.59	1.92 2.57					
19	4.38 8.18	3.52 5.93	3.13 5.01	2.90 4.50	2.74 4.17	2.63 3.94	2.55 3.77	2.48 3.63	2.43 3.52	2.38 3.43	2.34 3.36	2.31 3.30	2.26 3.19	2.21 3.12	2.15 3.00	2.11 2.92	2.07 2.84	2.03 2.76	2.00 2.70	1.96 2.63	1.94 2.60	1.91 2.54	1.89 2.51	1.88 2.49					
20	4.35 8.10	3.49 5.85	3.10 4.94	2.87 4.43	2.71 4.10	2.60 3.87	2.52 3.71	2.45 3.56	2.40 3.45	2.35 3.37	2.31 3.30	2.28 3.23	2.23 3.13	2.18 3.05	2.12 2.94	2.08 2.86	2.04 2.77	1.99 2.69	1.96 2.63	1.92 2.56	1.90 2.53	1.87 2.47	1.85 2.44	1.84 2.42					
21	4.32 8.02	3.47 5.78	3.07 4.87	2.84 4.37	2.68 4.04	2.57 3.81	2.49 3.65	2.42 3.51	2.37 3.40	2.32 3.31	2.28 3.24	2.25 3.17	2.20 3.07	2.15 2.99	2.09 2.88	2.05 2.80	2.00 2.72	1.96 2.63	1.93 2.58	1.89 2.51	1.87 2.47	1.84 2.42	1.82 2.38	1.81 2.36					
22	4.30 7.94	3.44 5.72	3.05 4.82	2.82 4.31	2.66 3.99	2.55 3.76	2.47 3.59	2.40 3.45	2.35 3.35	2.30 3.26	2.26 3.18	2.23 3.12	2.18 3.02	2.13 2.94	2.07 2.83	2.03 2.75	1.98 2.67	1.93 2.58	1.91 2.53	1.87 2.46	1.84 2.42	1.81 2.37	1.80 2.33	1.78 2.31					
23	4.28 7.88	3.42 5.66	3.03 4.76	2.80 4.26	2.64 3.94	2.53 3.71	2.45 3.54	2.38 3.41	2.32 3.30	2.28 3.21	2.24 3.14	2.20 3.07	2.14 2.97	2.10 2.89	2.04 2.78	2.00 2.70	1.96 2.62	1.91 2.53	1.88 2.43	1.84 2.41	1.82 2.37	1.79 2.33	1.77 2.28	1.76 2.26					
24	4.26 7.82	3.40 5.61	3.01 4.72	2.78 4.22	2.62 3.90	2.51 3.67	2.43 3.50	2.36 3.36	2.30 3.25	2.26 3.17	2.22 3.09	2.18 3.03	2.13 2.92	2.09 2.85	2.02 2.74	1.98 2.66	1.94 2.58	1.89 2.49	1.86 2.44	1.82 2.34	1.80 2.33	1.76 2.27	1.74 2.23	1.73 8.21					
25	4.24 7.77	3.38 5.57	2.99 4.68	2.76 4.18	2.60 3.86	2.49 2.63	2.41 3.44	2.34 3.32	2.28 3.21	2.24 3.13	2.20 3.05	2.16 2.90	2.11 2.89	2.06 2.81	2.00 2.70	1.96 2.62	1.92 2.54	1.87 2.45	1.84 2.40	1.80 2.32	1.77 2.29	1.74 2.23	1.72 2.19	1.71 2.17					
26	4.22 7.72	3.37 6.53	2.98 4.64	2.74 4.14	2.59 3.82	2.47 3.59	2.39 3.45	2.32 3.29	2.27 3.17	2.22 3.09	2.18 3.02	2.15 2.96	2.10 2.86	2.05 2.77	1.99 2.66	1.95 2.58	1.90 2.50	1.85 2.41	1.82 2.36	1.78 2.28	1.76 2.25	1.72 2.19	1.70 2.15	1.69 2.12					

Reprinted by permission from Statistical Methods by George W. Snedecor and William G. Cochran, 6th ed., (c) 1967, by the Iowa State University Press, Ames Iowa.

Table H. Critical values of F (continued).Degrees of
freedom
for lesser

Degrees of freedom for greater mean square

mean square	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
27	4.21 7.68	3.35 5.49	2.96 4.60	2.73 4.11	2.57 3.79	2.46 3.58	2.37 3.39	2.30 3.26	2.25 3.14	2.20 3.08	2.16 2.98	2.13 2.93	2.08 2.83	2.03 2.74	1.97 2.68	1.93 2.55	1.88 2.47	1.84 2.38	1.80 2.33	1.76 2.25	1.74 2.21	1.71 2.16	1.68 2.12	1.67 2.10
28	4.20 7.64	3.34 5.45	2.95 4.57	2.71 4.07	2.56 3.76	2.44 3.53	2.36 3.36	2.29 3.23	2.24 3.11	2.19 3.03	2.15 2.95	2.12 2.90	2.06 2.80	2.02 2.71	1.96 2.60	1.91 2.52	1.87 2.44	1.81 2.35	1.78 2.30	1.75 2.22	1.72 2.16	1.69 2.13	1.67 2.09	1.65 2.04
29	4.18 7.60	3.33 5.42	2.93 4.54	2.70 4.04	2.54 3.73	2.43 3.50	2.35 3.33	2.28 3.20	2.22 3.08	2.18 3.00	2.14 2.92	2.10 2.87	1.05 2.77	2.00 2.68	1.94 2.57	1.90 2.49	1.85 2.41	1.80 2.32	1.77 2.27	1.73 2.19	1.71 2.15	1.68 2.10	1.65 2.04	1.64 2.03
30	4.17 7.56	3.32 5.39	2.92 4.51	2.69 4.02	2.56 3.70	2.42 3.47	2.34 3.20	2.27 3.17	2.21 3.05	2.16 2.98	2.12 2.90	2.09 2.84	2.04 2.74	1.99 2.66	1.93 2.55	1.89 2.47	1.84 2.38	1.79 2.29	1.76 2.24	1.72 2.16	1.69 2.13	1.66 2.07	1.64 2.03	1.62 2.01
32	4.15 7.50	3.30 5.34	2.90 4.46	2.67 3.97	2.51 3.66	2.40 3.42	2.32 3.25	2.25 3.12	2.19 3.01	2.14 2.94	2.10 2.86	2.07 2.80	2.02 2.70	1.97 2.62	1.91 2.51	1.86 2.42	1.82 2.34	1.76 2.25	1.74 2.20	1.69 2.12	1.67 2.08	1.64 2.02	1.61 1.98	1.59 1.96
34	4.13 7.44	3.28 5.29	2.88 4.42	2.65 3.98	2.49 3.61	2.38 3.38	2.30 3.21	2.23 3.08	2.17 2.97	2.12 2.89	2.08 2.82	2.05 2.76	2.00 2.66	1.95 2.58	1.89 2.47	1.84 2.38	1.80 2.30	1.74 2.21	1.71 2.15	1.67 2.03	1.64 2.04	1.61 1.98	1.59 1.94	1.57 1.91
36	4.11 7.39	3.26 5.25	2.86 4.33	2.63 3.89	2.48 3.58	2.36 3.35	2.28 3.18	2.21 3.04	2.15 2.94	2.10 2.86	2.06 2.73	2.03 2.72	1.98 2.62	1.93 2.54	1.87 2.42	1.82 2.35	1.78 2.26	1.72 2.17	1.69 2.12	1.65 2.04	1.62 2.00	1.59 1.94	1.56 1.90	1.55 1.87
38	4.10 7.35	3.25 5.21	2.85 4.34	2.62 3.86	2.46 3.54	2.35 3.32	2.26 3.15	2.19 3.02	2.14 2.91	2.09 2.82	2.05 2.75	2.02 2.69	1.96 2.59	1.92 2.51	1.85 2.40	1.80 2.32	1.76 2.22	1.71 2.14	1.67 2.03	1.63 2.00	1.60 1.97	1.57 1.90	1.54 1.85	1.53 1.84
40	4.08 7.31	3.23 5.18	2.84 4.31	2.61 3.83	2.45 3.51	2.34 3.29	2.25 3.12	2.18 2.99	2.12 2.88	2.07 2.80	2.04 2.73	2.00 2.64	1.95 2.54	1.90 2.49	1.84 2.27	1.79 2.29	1.74 2.20	1.69 2.11	1.66 2.05	1.61 1.97	1.59 1.94	1.55 1.83	1.53 1.54	1.51
42	4.07 7.27	3.22 5.15	2.83 4.29	2.59 3.80	2.44 3.49	2.32 3.26	2.24 3.10	2.17 2.96	2.11 2.86	2.06 2.77	2.02 2.70	1.99 2.64	1.94 2.54	1.89 2.46	1.82 2.35	1.78 2.26	1.73 2.17	1.68 2.06	1.64 2.02	1.60 1.94	1.57 1.91	1.54 1.85	1.51 1.80	1.49 1.78
44	4.06 7.24	3.21 5.12	2.82 4.26	2.58 3.73	2.43 3.46	2.31 3.24	2.23 3.07	2.16 2.94	2.10 2.84	2.05 2.75	2.01 2.68	1.98 2.62	1.92 2.52	1.88 2.44	1.81 2.32	1.76 2.24	1.72 2.15	1.65 2.04	1.63 2.00	1.59 1.92	1.56 1.88	1.52 1.82	1.50 1.76	1.43 1.75
46	4.05 7.21	3.20 5.10	2.81 4.24	2.57 3.76	2.42 3.44	2.30 3.22	2.22 3.05	2.14 2.92	2.09 2.82	2.04 2.73	2.00 2.65	1.97 2.60	1.91 2.50	1.87 2.42	1.80 2.30	1.75 2.22	1.71 2.13	1.65 2.04	1.62 1.98	1.57 1.90	1.54 1.86	1.51 1.80	1.48 1.76	1.46 1.72
48	4.04 7.19	3.19 5.08	2.80 4.22	2.56 3.74	2.41 3.42	2.30 3.20	2.21 3.04	2.14 2.90	2.08 2.80	2.03 2.71	1.99 2.64	1.96 2.58	1.90 2.48	1.86 2.40	1.79 2.28	1.74 2.20	1.70 2.11	1.64 2.02	1.61 1.96	1.56 1.88	1.53 1.84	1.50 1.78	1.47 1.72	1.45 1.70
50	4.03 7.17	3.18 5.06	2.79 4.20	2.56 3.72	2.40 3.41	2.29 3.18	2.20 3.02	2.13 2.88	2.07 2.78	2.02 2.70	1.98 2.62	1.95 2.56	1.90 2.46	1.85 2.39	1.78 2.26	1.74 2.18	1.69 2.10	1.63 2.00	1.60 1.94	1.55 1.86	1.52 1.82	1.48 1.76	1.46 1.71	1.44 1.65
55	4.02 7.12	3.17 5.01	2.78 4.16	2.54 3.68	2.38 3.37	2.27 3.15	2.18 2.98	2.11 2.85	2.05 2.75	2.00 2.66	1.97 2.59	1.93 2.53	1.88 2.43	1.83 2.35	1.76 2.23	1.72 2.15	1.67 2.06	1.61 1.96	1.58 1.90	1.52 1.82	1.50 1.78	1.46 1.71	1.43 1.66	1.41 1.64
60	4.00 7.08	3.15 4.98	2.76 4.13	2.52 3.65	2.37 3.34	2.25 3.12	2.17 2.95	2.10 2.82	2.04 2.72	1.99 2.63	1.95 2.56	1.92 2.50	1.86 2.40	1.81 2.32	1.75 2.20	1.70 2.12	1.65 2.03	1.59 1.93	1.56 1.87	1.50 1.79	1.48 1.74	1.44 1.68	1.41 1.63	1.39 1.60
65	3.99 7.04	3.14 4.95	2.75 4.10	2.51 3.62	2.36 3.31	2.24 3.09	2.15 2.93	2.08 2.79	2.02 2.70	1.98 2.61	1.94 2.54	1.90 2.47	1.85 2.37	1.80 2.30	1.73 2.18	1.68 2.09	1.63 2.00	1.57 1.90	1.54 1.84	1.49 1.76	1.46 1.71	1.42 1.64	1.39 1.60	1.37 1.56
70	3.98 7.01	3.13 4.92	2.74 4.08	2.50 3.60	2.35 3.29	2.23 3.07	2.14 2.91	2.07 2.77	2.01 2.67	1.97 2.59	1.93 2.51	1.89 2.45	1.84 2.35	1.79 2.28	1.72 2.15	1.67 2.07	1.62 1.98	1.56 1.88	1.53 1.82	1.47 1.74	1.45 1.69	1.40 1.62	1.37 1.56	1.35 1.53
80	3.96 6.96	3.11 4.83	2.72 4.04	2.48 3.56	2.33 3.25	2.21 3.04	2.12 2.67	2.05 2.74	1.99 2.64	1.95 2.55	1.91 2.48	1.88 2.41	1.82 2.32	1.77 2.24	1.70 2.11	1.65 2.03	1.60 1.94	1.54 1.84	1.51 1.78	1.45 1.70	1.42 1.65	1.38 1.57	1.35 1.52	1.32 1.49
100	3.94 6.90	3.09 4.82	2.70 3.98	2.46 3.51	2.30 3.20	2.19 2.98	2.10 2.82	2.03 2.69	1.97 2.59	1.92 2.51	1.88 2.43	1.85 2.36	1.79 2.26	1.75 2.19	1.68 2.06	1.63 1.98	1.57 1.89	1.51 1.79	1.48 1.73	1.42 1.64	1.39 1.59	1.34 1.51	1.30 1.45	1.28 1.43
125	3.92 6.84	3.07 4.78	2.68 3.94	2.44 3.47	2.29 3.17	2.17 2.95	2.08 2.79	2.01 2.65	1.95 2.56	1.90 2.47	1.86 2.40	1.83 2.33	1.77 2.23	1.72 2.15	1.65 2.03	1.60 1.94	1.55 1.85	1.49 1.75	1.45 1.68	1.39 1.59	1.36 1.54	1.31 1.46	1.27 1.40	1.25 1.37
150	3.91 6.81	3.06 4.75	2.67 3.91	2.43 3.44	2.27 3.14	2.16 2.92	2.07 2.76	2.00 2.62	1.94 2.53	1.89 2.44	1.85 2.37	1.82 2.30	1.76 2.20	1.71 2.12	1.64 2.00	1.59 1.91	1.54 1.83	1.47 1.72	1.44 1.68	1.37 1.56	1.34 1.51	1.29 1.42	1.25 1.37	1.22 1.33
200	3.89 6.76	3.04 4.71	2.65 3.58	2.41 3.41	2.26 3.11	2.14 2.90	2.05 2.73	1.98 2.60	1.92 2.50	1.87 2.41	1.83 2.34	1.80 2.28	1.74 2.17	1.69 2.09	1.62 1.97	1.57 1.88	1.52 1.79	1.45 1.69	1.42 1.62	1.35 1.53	1.32 1.48	1.26 1.39	1.22 1.32	1.19 1.30
400	3.86 6.70	3.02 4.66	2.62 3.83	2.39 3.36	2.23 3.06	2.12 2.85	2.03 2.69	1.96 2.55	1.90 2.46	1.85 2.37	1.81 2.29	1.78 2.23	1.72 2.12	1.67 2.04	1.60 1.92	1.54 1.84	1.49 1.74	1.42 1.64	1.38 1.57	1.32 1.47	1.28 1.42	1.22 1.32	1.16 1.24	1.13 1.19
1000	3.85 6.68	3.00 4.62	2.61 3.80	2.38 3.34	2.22 3.04	2.10 2.82	2.02 2.66	1.95 2.53	1.98 2.43	1.84 2.34	1.80 2.26	1.76 2.20	1.70 2.09	1.65 2.01	1.58 1.89	1.53 1.81	1.47 1.71	1.41 1.61	1.36 1.54	1.30 1.44	1.26 1.38	1.19 1.28	1.13 1.19	1.08 1.11
∞	3.84 6.64	2.99 4.60	2.60 3.78	2.37 3.32	2.21 3.02	2.09 2.80	2.01 2.64	1.94 2.31	1.88 2.41	1.83 2.32	1.79 2.24	1.75 2.18	1.69 2.07	1.64 1.99	1.57 1.87	1.52 1.79	1.46 1.69	1.40 1.62	1.35 1.52	1.28 1.41	1.24 1.36	1.17 1.25	1.11 1.15	1.00 1.00

INDEX

- a-posteriori* tests 301-302
- a-priori* tests 300-302, 304-305, 311
- acceptance region 113-115
- age scores 272
- alternate-forms method 259-260
- alternative hypothesis (H_a) 111-115
- analysis of variance 287-328
 - assumptions 289-290
 - models 288-289, 294, 299-300, 326
 - nonparametric 310-315, 340
 - one-way anova 288, 291-300, 303-309, 326, 341
 - two-way anova 288, 315-326
- assumptions 118
- bar diagram 20-22
- Bernoulli distribution 101, 104-105
- Binet-Simon scale 272
- binomial distribution 100-105, 107, 216, 221
- biserial r 183-186, 254
- bivariate distribution 156
- bivariate statistics 144
- Bonferroni method 301, 305
- C* scale 275-276
- central moments 67
- chi square tests 208-221, 335
- classification variables 5, 246, 281-282, 289, 291
- clumped distribution 71
- coefficient
 - of dispersion 71, 102-103, 106-107
 - of equivalence 259-260
 - of internal consistency 261
 - of quartile deviation 71
 - of reliability 257
 - of stability 259
 - of stability and equivalence 260
 - of validity 264
 - of variation 70
- composite rank test 231-237
- concurrent validity 267
- confidence interval 97, 99-100
- construct validity 266
- content validity 265-266
- contingency coefficient 191-193
- convergent validity 266, 268
- correlation 12, 144-145
 - linear 145-162, 287, 336-337
 - multiple 164-167, 287
 - nonparametric 167-178, 187-189, 191-193
 - partial 162-166, 287
 - table 156-159, 268
- criterion-related validity 266-267
- critical region 113, 115-117
- critical scores 88-89, 97, 113-114, 123-124, 126
- cumulative frequency 30
- decile 47
- degrees of freedom 82, 123
- dependent variables 4-5, 247, 281-285, 287-288
- descriptive statistics 11-12
- deviation IQ 277-278
- difference between means 110-111
 - standard error 74, 78-79
 - tests for significance 111, 117-143
- difficulty value 249-253
- discriminant validity 266, 268
- discriminatory value 249, 253-256
- equivalent-forms reliability 259-260
- equivalent groups 139-140, 284
- error
 - of inference 114-115
 - of measurement 257
 - of prediction 196
- error variance 292, 317
- expectancy table 268
- experimental design 281-288
- exponential curve 37-38
- F* ratio 166-167, 288, 291-294, 296, 298-305, 307-309, 317-319, 321-323, 325, 327, 341
- F* test, Scheffe's 301-302, 304
- factor analysis 268-270
- factor loading 270
- factor theory 269-270
- factor validity 270
- factorial experiment 285

- Fisher's z 147-148, 155
- frequency
 - cumulative 30
 - distribution 15-19, 22-30, 156
 - curve 30
 - polygon 22-26
- G test 222-227
- Gabriel's $SS-STP$ 302
- geometric mean 45-46
- histogram 26-30
- homoscedasticity 118, 123, 290
- independent group experiment 118-132, 283-284, 330-331, 333
- independent variables 5-6, 247-248, 281-285
- index of discrimination 254-256
- interval scale 3
- intervening variables 249
- intrinsic validity 264
- item analysis 249-256
- item-criterion correlation 184, 187, 189, 254
- item-item correlation 187, 254
- item-total correlation 179, 184, 253-254
- item validity 253
- Kendall's tau 172-178, 287
- Kuder-Richardson formulae 261-263
- Kruskal-Wallis test 310-315, 340
- kurtosis 94-95
- linear graph 34-37
- linear transformation 80-81, 271
- Mann-Whitney U test 237-241, 311-313, 334, 340
- matched-pair groups 139-140, 284
- mean 41-45
- mean deviation 57
- median 49-53
- median test 242-245
- mental age 272
- mode 53-55
- multiple comparison 300-302, 304-305
- multiple correlation 164-167, 287
- multivariate statistics 144, 162, 164
- nonlinear transformations 81, 271, 290
- norm 270-278
- normal distribution 86-92, 98
 - best-fitting 89-92
- normalized standard scores 273
- null hypothesis (H_0) 111-116
- ogive 30-33
- omega square 294, 299-300, 305, 341
- one-tail test 88, 116-117, 124-125, 128-130, 132, 137, 139, 142-143
- one-way anova 288, 291-300, 303-309, 326, 341
- parallel-forms method 259-260
- parameter 11
- partial correlation 162-164
- part-whole correlation 258
- percentile 46-48
 - rank 48-49
 - scale 271-272
- phi coefficient 187-189, 254-256
- pie diagram 20
- point-biserial r 179-183, 253-254
- Poisson distribution 105-108, 216
- population 6-7
- power test 250
- prediction statistics 12, 193
- predictive validity 266-267
- probability 84-85
 - distributions 85-108
- product-moment r 133, 139-142, 145-162, 287, 329, 336-337
- prospective method 282-283
- quantiles 41, 46-48
- quartile 47, 67-71
 - deviation 67-71
- randomization 7-11, 283-287, 289-290
- randomized block 284-285
- range 56-57
- ratio IQ 272-273
- ratio scale 3
- rational equivalence 261
- regression 12, 193-195
 - models 194-195
 - multiple 194, 203-206
 - simple 195-203, 338-339
- rejection region 113, 115-117
- relevant validity 264

relevant variables 6, 248, 281-283, 287
 reliability 249, 257-263, 267-268, 287-288
 repulsed distribution 71
 retrospective method 281-282

sample 7
 size 285

sampling 7-11, 286-287
 sampling distributions 73-74
 sampling errors 73, 110
 sampling statistics 12, 73-83
 scattergram 33-34
 semi-logarithmic graph 38-39
 signed rank test 227-231
 significance level 112-114
 significance test 110-111, 117-143, 329-334
 single-group experiment 132-143, 284, 332
 skewness 92-94

Spearman-Brown formulae 260-261

Spearman's rho 167-171, 287

speed test 250

split-half reliability 260-261

standard deviation 57-64, 70

standard error 74-80

of differences 74, 78-80, 110

of estimate 196, 201, 204-206, 264

of mean 74-79

of measurement 258

of proportion 78-80

standard score 80-81, 273

Stanford-Binet scale 272-273, 278

stanine 276-277

statistics 1

of dispersion 12, 56-72

of location 11-12, 41-55

nonparametric 208

prediction 12

sampling 12, 73-83

sum of squares

between-columns 316, 318

between-groups 291

between-rows 316, 318

critical 302

interaction 316-317

partitioning 291-292, 294-297, 299-300, 303-307,
 309, 315-319, 341

sum of squares (contd.)

residual 317, 319

total 291, 315-316

within-groups 292

t distribution 95-98

T scale 273-275

t score 95-100, 123-143, 301, 304-305, 329-332, 336-337

T statistic 228-229, 231-237

t tests 123-143, 301, 304-305, 329-332, 336-337

multiple comparison 301, 304-305

test construction 249-250

test-retest reliability 259

tetrachoric *r* 189-191, 254

treatment variables 5, 248, 281-282, 287-289, 294,
 297, 303, 305

truncated distributions 184

two-tail test 88-89, 115-122, 124, 126-129, 131-137,
 140-142

two-way anova 288, 315-326

unit normal curve 86-92

validity 249, 263-268, 287-288

variables 2-6, 247-249, 281-285, 287

variance 64-67, 71

added treatment component 289

added variance component 289, 294, 296, 308

between-columns 316, 318

between-groups 292

between-rows 316, 318

error 292, 317

interaction 316-317, 319

ratio (see *F* ratio)

residual 317

total 292, 316

within-cells 317

within-groups 292

Wechsler intelligence scale 277

Wilcoxon's rank tests 227-237, 287

Yates' correction 210, 214-216, 222, 225, 227

z score 80-81, 86-92, 97-100, 117-123, 185-186, 238,
 240-241, 313-314, 333-334